

Trabajo Fin de Grado

Estudio de Sistemas Avanzados de Asistencia al Conductor (ADAS) en vehículos y propuesta de aplicación de técnicas de seguimiento de la mirada para su mejora.

Study of Advanced Driver Assistance Systems (ADAS) in vehicles and a proposal for the application of eye-tracking techniques.

Autor/es

Blanca Lasheras Hernández

Director/es

Daniel Martín Serrano

Ponente

Belén Masiá Corcoy

Escuela de Ingeniería y Arquitectura

2021

Resumen

Durante las últimas décadas, el campo de la automoción ha experimentado un notable desarrollo tecnológico. En gran medida, esto se debe al ambicioso objetivo de la creación de un vehículo autónomo, en el cual el conductor es relevado de sus tareas. No obstante, dado el nivel de desarrollo actual, existen ciertas tareas en las cuales el conductor sigue siendo imprescindible.

Muchas de ellas tienen una repercusión directa sobre la seguridad vial, que juega un papel fundamental en la seguridad de los usuarios durante su desplazamiento por las vías. Para garantizarla, es crucial que el conductor esté alerta y pueda responder ante situaciones imprevistas que puedan entrañar un posible riesgo. En este contexto, surgen los Sistemas Avanzados de Asistencia al Conductor (en inglés, *Advanced Driver Assistance Systems*, ADAS), que apoyan la tarea de conducción, asistiendo al conductor en la toma de decisiones, o incluso asumiendo el control de la conducción parcialmente si la situación lo requiere, disminuyendo así el número de accidentes que se producen.

En este contexto, uno de los objetivos principales de este proyecto es el estudio del estado del arte de los ADAS empleados en la actualidad, a través de su clasificación mediante una taxonomía propuesta en la literatura, así como un análisis de su funcionamiento. Dentro de este amplio ámbito, se hace especial énfasis en ADAS basados en *visión por computador*, que tienen la capacidad de monitorizar tanto el exterior del vehículo como al propio conductor. Además, se estudian los ADAS que incluyen seguimiento de la mirada, puesto que proporcionan información del estado de atención del conductor, que es crítico en la garantía de la seguridad vial.

En el marco de los ADAS basados en visión y seguimiento de mirada, y en base a las limitaciones de los sistemas actuales, se propone también un sistema basado en seguimiento de mirada que permite inferir información sobre el estado de atención del conductor a partir de información visual del exterior del vehículo. Para ello, se estudian modelos de aprendizaje profundo capaces de inferir las regiones del entorno visual del conductor a las que es más probable que este mire. Adicionalmente, se estudian las limitaciones de tal modelo, y se proponen diversas mejoras, que resultan en un sistema más preciso, incluyendo la capacidad de generalización en un mayor número de situaciones, así como su funcionamiento con diferentes condiciones meteorológicas y lumínicas. El sistema desarrollado es versátil, y potencialmente implementable en un ADAS de bajo coste.

Agradecimientos

En primer lugar, quiero agradecer a mis directores, Belén y Dani, el constante apoyo que han supuesto para mí a lo largo de todo el proyecto. Con su guía, ayuda, buen hacer y disposición en todo momento, he aprendido muchas cosas nuevas, y descubierto nuevas fronteras que ni siquiera imaginaba. Gracias por enseñarme también que, a pesar de las adversidades, hay que seguir intentándolo.

También, agradecer a todos los miembros del grupo *Graphics and Imaging Lab* de la Universidad de Zaragoza su acogida y buen ambiente desde el principio, interesándose por mi trabajo, aportando nuevos puntos de vista y regalándome sus conocimientos.

A Pablo, que ha estado a mi lado apoyándome desde antes de saber que sería Ingeniera, y durante toda la carrera, escuchándome y ayudándome en las situaciones más difíciles, y celebrando los buenos momentos. Haces de cada instante una aventura emocionante.

A mis abuelos, Paco y Luchi, y a mis yayos, Pilar y Aparicio, que me enseñan a ser mejor persona a través de sus historias y valores. *Paciencia, trabajo y tranquilidad*. A mis tíos, tías, primos, primas, y demás familia. Siempre os tengo presentes.

A mis padres, Susana y Javier, que me dejan recortes de periódico de noticias relacionadas con lo que hago en la Universidad o en el Conservatorio, me apoyan, se interesan por todo lo que hago, y me escuchan y aconsejan en todo momento.

A mi hermano, Dani, que aunque ya no esté en el cuarto de al lado, hace todo por estar cerca siempre.

A Rebeca, que siempre me apoya en las grandes decisiones y me anima a afrontar nuevos retos. Gracias por hacer de estratega y querer siempre lo mejor para mí. Sin ti, no hubiera sido posible.

A mis amigos y amigas de la Universidad, el Conservatorio, Viena, y a todas aquellas personas que, de una forma u otra, han estado conmigo a lo largo del proyecto y de la carrera, sois un gran apoyo.

Gracias.

Índice

1. Introducción	12
1.1. Contexto del proyecto	12
1.2. Objetivos y alcance del proyecto	13
1.3. Planificación y herramientas	14
2. Sistemas Avanzados de Asistencia al Conductor (ADAS)	16
2.1. Vehículo autónomo	16
2.1.1. Niveles de autonomía (<i>SAE</i>)	16
2.1.2. Otras clasificaciones de niveles de autonomía	18
2.1.3. Nivel de autonomía actual	19
2.2. ADAS en vehículos	19
2.2.1. Tipos de ADAS	20
2.2.2. Sensores	20
2.2.3. Flujo de datos de un ADAS	22
2.2.4. Sistemas de alerta/respuesta	23
3. ADAS basados en visión	25
3.1. Captura de información	25
3.1.1. Sensores	25
3.1.2. Monitorización	26
3.2. Flujo de datos	28
3.3. Respuesta del vehículo	30
3.4. Seguimiento de la mirada en ADAS	32
3.4.1. Estado y comportamiento del conductor	32
3.4.2. Técnicas de seguimiento de la mirada	34
3.4.3. Algoritmos de procesamiento de la mirada en ADAS	35
3.5. Limitaciones	36
4. Predicción de la atención en conducción	39
4.1. Motivación y contexto	39
4.1.1. Técnicas de <i>deep learning</i>	40
4.1.2. Saliencia	41
4.2. Desarrollo de un modelo genérico de predicción de saliencia	42
4.2.1. Segmentación semántica	45
4.3. Desarrollo de modelos <i>específicos</i> de predicción de saliencia	47
4.4. Resumen	49

5. Conclusiones	52
5.1. Trabajo futuro	53
5.2. Nota personal	53
Bibliografía	55
A. Deep learning	64
A.1. Redes neuronales artificiales	64
A.1.1. Etapas de desarrollo de una red neuronal	64
A.2. Redes neuronales convolucionales	67
A.3. Modelo <i>encoder-decoder</i>	69
B. Conjunto de datos de <i>Dr(eye)ve</i>	70
B.1. Datos empleados (<i>Dataset</i>)	70
B.1.1. Datos empleados en el proyecto	70
B.1.2. Data augmentation	72

Índice de tablas

2.1. Niveles de autonomía de la conducción según SAE Internacional.	17
2.2. Correspondencia entre los niveles de automatización SAE, BAST y NHTSA. . . .	18
3.1. Resumen de las técnicas de medición empleadas en la detección de la distracción del conductor en el vehículo.	35
3.3. Resumen de los algoritmos de <i>eye tracking</i> más empleados.	37
4.1. Estudio de ablación de la incorporación de la segmentación semántica, así como de la ponderación de la función de pérdida KL.	46
4.2. Estudio de ablación que permite identificar el <i>overfitting</i> de la red.	47
4.3. Estudio de ablación en el que se comparan las métricas obtenidas con modelos entrenados en condiciones específicas, con los resultados de un modelo entrenado genéricamente. En dos de los tres casos, el modelo específico genera mejores cifras.	48
4.4. Estudio de ablación en el que se analizan los resultados para el caso diurno. Se observa una mejora de las métricas al emplear aumento de datos, superando los valores obtenidos con anterioridad.	49
B.2. Resumen de las características del conjunto de datos de <i>Dr(eye)ve</i>	71
B.3. Transformaciones aplicadas a los datos en condiciones diurnas, obtenidas de la librería <i>Albumentations</i>	72

Índice de figuras

1.1. Ejemplos de mapas de saliencia predichos por el modelo propuesto en este proyecto.	13
1.2. Diagrama de Gantt con las horas dedicadas a cada tarea del proyecto.	15
2.1. Taxonomía de un ADAS en función de los tipos de sensores que emplea.	20
2.2. Disposición frecuente de sensores en el vehículo de tal forma que la mayor parte de su entorno quede cubierta.	21
2.3. Esquema con los principales módulos que frecuentemente conforman un ADAS genérico.	22
2.4. Diagrama general del flujo de datos de un ADAS desde su obtención hasta la respuesta del sistema.	23
3.1. Disposición usual de las cámaras en el vehículo.	26
3.2. Procedimiento de reconocimiento de señales mediante segmentación.	27
3.3. Detección de bordes de imagen mediante algoritmo de Canny e identificación de los límites de carril con la transformada de Hough.	28
3.4. Detección y seguimiento de objetos y vehículos, y cálculo de distancias como parte del sistema de evasión de colisiones.	29
3.5. Flujo de datos con seis etapas en ADAS basados en visión.	30
3.6. Aspectos determinantes para evaluar el nivel de fatiga del conductor.	31
3.7. Etapas de procesamiento de datos de distinta naturaleza en ADAS que monitorizan la mirada del conductor.	33
3.8. Representación genérica de las técnicas empleadas para el seguimiento del ojo y las coordenadas de la mirada.	34
3.9. Diagrama de flujo de un algoritmo de <i>eye tracking</i>	36
4.1. Redes neuronales: estructura y terminología general.	40

4.2.	Mapas de saliencia con información de la mirada del conductor obtenidos mediante la superposición de 25 fotogramas de secuencia de conducción.	42
4.3.	Esquema del modelo <i>encoder-decoder</i> empleado inicialmente, formado por capas convolucionales.	43
4.4.	Esquema neuronal por capas del <i>encoder</i> y <i>decoder</i>	44
4.5.	Clasificación de los distintos elementos de la imagen mediante segmentación semántica en las ciudades alemanas Düsseldorf (izquierda) y Dortmund (derecha).	45
4.6.	Esquema del modelo <i>encoder-decoder</i> de doble rama, en el cual se incluye un modelo de segmentación semántica pre-entrenado.	46
4.7.	Ejemplo de las transformaciones realizadas sobre las imágenes del <i>dataset</i> , obtenidas de la librería <i>Albumentations</i>	48
4.8.	Comparativa de resultados cualitativos entre el modelo genérico y el modelo específico.	49
4.9.	Resultados obtenidos con los modelos optimizados en base a las tres condiciones lumínicas y meteorológicas planteadas.	51
A.1.	Redes neuronales: analogía entre el modelo biológico y el artificial.	65
A.2.	Representación de aplicación de un filtro a una imagen, entendida como una matriz de píxeles.	68
A.3.	Esquema de una capa convolucional. En este proyecto, formada por un único bloque convolucional: convolución, normalización y activación.	69
B.1.	Ejemplo del contenido del conjunto de datos de <i>Dr(eye)ve</i> . De izquierda a derecha: vista desde salpicadero, vista del conductor, mapa de saliencia y ruta georreferenciada.	71

Lista de Acrónimos

ADAS	<i>Advanced Driver Assistance System</i> (Sistema Avanzado de Asistencia al Conductor)
ADS	<i>Automated Driving System</i> (Sistema de Conducción Autónoma)
BASf	Instituto Federal Alemán de Investigación de Carreteras
CC	<i>Linear correlation coefficient</i> (Coeficiente de correlación lineal)
CNN	<i>Convolutional Neural Network</i> (Red neuronal convolucional)
EDA	<i>Emergency Driver Assistant</i> (Conductor asistente de emergencia)
fps	<i>Frames per second</i> (Fotogramas por segundo)
GPS	<i>Global positioning system</i> (Sistema de posicionamiento global)
HMI	<i>Human-Machine Interface</i> (Interacción persona-máquina)
IA	Inteligencia artificial
IMU	<i>Inertial measurement unit</i> (Unidad de medición inercial)
IR	<i>Infrared</i> (Infrarrojo)
KBS	<i>Knowledge-based Systems</i> (Sistemas basados en conocimiento)
KL	<i>Kullback-Leibler Divergence</i> (Divergencia de Kullback-Leibler)
LIDAR	<i>Laser Imaging Detection and Ranging</i> (Detección y Rango de Imágenes Láser)
NHTSA	<i>National Highway Traffic Safety Administration</i> (Administración Nacional de Seguridad del Tráfico en las Carreteras)
NN	<i>Artificial Neural Networks</i> (Redes neuronales artificiales)
NSS	<i>Normalized scanpath saliency</i> (Saliencia de la ruta de exoíración normalizada)
PERCLOS	<i>Percentage of eye closure</i> (Porcentaje de cierre ocular)
PMD	<i>Photonic mixer device</i> (mezclador fotónico)
PRC	<i>Percent road center</i> (Porcentaje del centro de carretera)

RCTA	<i>Rear Cross Traffic Alert</i> (Sistema de alerta de tráfico cruzado)
RGB	<i>Red Green Blue</i> (Rojo Verde Azul)
SAE	<i>Society of Automotive Engineers</i> (Sociedad de Ingenieros de la Automoción)
sfm	<i>structure-from-motion</i> (geometría epipolar)
SIM	<i>Similarity score</i> (Valor de similitud)

1. Introducción

1.1. Contexto del proyecto

En las últimas décadas, el desarrollo de una sociedad basada en el consumo en la que la demanda de bienes se produce de forma instantánea, el avance tecnológico en el ámbito de la conducción se está produciendo a un ritmo frenético. Además, los objetivos globales para evitar el cambio climático, establecidos en el Acuerdo de París, establecen nuevos requerimientos que llevan a los fabricantes de vehículos a reemplazar los motores de combustión interna por su alternativa eléctrica. Por todo ello, la industria automovilística se encuentra cada vez más orientada hacia la producción, en crecimiento, de vehículos eléctricos que, en un futuro, podrían operar de forma autónoma. Es en este contexto en el que surge el concepto de *vehículo autónomo*, con el objetivo de crear un automóvil totalmente automatizado en el que la única intervención del conductor sea para la selección del destino al comienzo del desplazamiento.

En el proceso por conseguir tal objetivo, se producen avances que afectan progresivamente al grado o nivel de automatización de los vehículos a través del uso creciente de sistemas automáticos, que requieren de la incorporación de grandes cantidades de componentes electrónicos (especialmente, microprocesadores). Hace una década, los vehículos de gama alta incorporaban en torno a 100 microprocesadores que funcionaban como unidades de control, ejecutando más de 100 millones de líneas de código durante su operación [1]. No obstante, hoy en día, la mayoría de vehículos ya cuentan con una cifra superior de componentes electrónicos integrados.

Actualmente, el desarrollo tecnológico hacia el vehículo autónomo se encuentra en un estado intermedio, gracias al empleo de Sistemas Avanzados de Asistencia al Conductor (en inglés, *Advanced Driver Assistance Systems*, ADAS) en vehículos. Su constante desarrollo y mejora a lo largo de las últimas décadas se debe al creciente interés por crear sistemas que, mediante la monitorización del entorno, proporcionen asistencia al conductor, facilitando sus procesos de toma de decisiones. Sin embargo, el empleo de ADAS que implican la autonomía parcial del vehículo ha de producirse de forma controlada, puesto que el conductor todavía no puede ser reemplazado en circunstancias que entrañan riesgos no abarcables por el sistema, pudiendo producirse accidentes con graves consecuencias [2, 3].

Los avances en la capacidad de computación actual, y el desarrollo de algoritmos más precisos y robustos juegan un papel fundamental en la creación de ADAS más complejos. Entre las herramientas más presentes en estos sistemas en los últimos años, destacan las basadas en el denominado *aprendizaje profundo* (en inglés, *deep learning*), que permite a los ordenadores llevar a cabo ciertas tareas que, hasta el momento, podían ser llevadas a cabo únicamente por el ser



Figura 1.1: Ejemplos de predicción de saliencia con el modelo implementado en este trabajo de fin de grado para dos escenas de conducción distintas. *Izquierda*: Imagen RGB que recibe el modelo. Muchos de los ADAS actuales están basados en visión, y, por tanto, el coste de obtención de una imagen del entorno del vehículo es mínimo. *Centro*: Mapa de saliencia real para la escena en cuestión, obtenido a partir de la agregación de datos empíricos de conductores. Las regiones blancas indican qué partes de la escena son más propensas a atraer la mirada de un conductor. *Derecha*: Predicción de un mapa de saliencia con el modelo desarrollado en este proyecto. Aprovechando el conocimiento del comportamiento visual de conductores, se ha desarrollado un modelo que, tomando una única imagen RGB (izquierda) como entrada, es capaz de predecir qué zonas del entorno son más propensas a atraer la mirada del conductor. El modelo ofrece unos resultados precisos y próximos a los datos empíricos (centro).

humano, como la asistencia sanitaria, la traducción de textos e imágenes, el reconocimiento facial y visual, la generación de textos, o la conducción autónoma. En este último campo, el aprendizaje profundo se emplea para identificar objetos o elementos del entorno, estimar posiciones de los mismos, o incluso analizar el estado del conductor. Este último es un caso crítico, puesto que tiene gran impacto sobre la seguridad vial: Factores como la distracción o el estado de somnolencia del conductor pueden ser causantes de accidentes fatales. Para obtener información del conductor, se emplean con frecuencia técnicas basadas en seguimiento de mirada (en inglés, *eye tracking*), que permiten conocer la dirección de la mirada del conductor, o patrones de comportamiento visual, mediante la interpretación precisa de las posiciones y movimientos oculares a través de la pupila o el centro del iris, proporcionando información que puede repercutir enormemente en la seguridad vial.

1.2. Objetivos y alcance del proyecto

En este proyecto fin de grado, se lleva a cabo un estudio del estado del arte en ADAS. Previo a ello, se va a estudiar la situación actual en el desarrollo del vehículo autónomo, y se va a ahondar en una clasificación de distintos ADAS actuales basándose en una taxonomía existente [4] donde se categorizan en función de sus sensores. Adicionalmente, se va a analizar el funcionamiento de los mismos, incluyendo el flujo de datos y los sistemas de alerta y respuesta que los conforman.

Particularmente, y dada su importancia en la monitorización del entorno, y por ende en la seguridad vial, se profundiza en ADAS basados en visión, analizando las diferentes etapas de procesamiento de datos que lo componen. De entre estos, se comenta cómo el seguimiento de la

mirada juega un papel fundamental en los algoritmos que estos sistemas emplean.

Finalmente, y con todo lo aprendido en estos estudios, se desarrolla un sistema basado en aprendizaje profundo, capaz de inferir qué regiones del entorno son más propensas a atraer la mirada del conductor (ver Figura 1.1). Para esto, se hace uso de la saliencia como principal indicador de la atención humana. A lo largo de dicho proceso, se estudian las arquitecturas más convenientes para trabajar en problemas de tratamiento de imágenes, así como en trabajos relacionados con la atención humana, y funciones de pérdida y métricas relevantes para el problema a abordar. Además, se analizan las limitaciones o carencias que surgen de estas propuestas, y se proponen diferentes técnicas para paliarlas, como el uso de aumento de datos, la simplificación de la arquitectura del modelo, o la creación de modelos específicos para diferentes condiciones durante la conducción.

El objetivo general de este trabajo es el estudio y análisis de los ADAS actuales, así como la propuesta de aplicación de técnicas de seguimiento de mirada para la asistencia al conductor. Para ello, se establecen los siguientes objetivos específicos:

- Estudio del nivel de desarrollo actual en la conducción autónoma (Sección 2.1).
- Estudio del estado del arte en ADAS en vehículos. Clasificación y análisis del funcionamiento de los mismos (Sección 2.2).
- Estudio de ADAS basados en visión, incluyendo su funcionamiento y características principales (Secciones 3.1, 3.2 y 3.3).
- Estudio de técnicas de seguimiento de la mirada en ADAS (Sección 3.4).
- Diseño, implementación y evaluación de varios modelos de predicción de la atención en la conducción (Sección 4).

Complementariamente, al final del documento se encuentran los Anexos, en los cuales se proporciona una explicación con mayor detalle sobre el aprendizaje profundo (Anexo A), y sobre los datos empleados en el proyecto (Anexo B).

Este proyecto se lleva a cabo en el grupo de investigación *Graphics and Imaging Lab*, en la Universidad de Zaragoza. El trabajo del grupo se centra en los *gráficos por computador*, realizándose investigación en áreas de renderizado físicamente correcto, procesamiento de imágenes, fotografía computacional, realidad virtual, o percepción aplicada, entre otros. El trabajo del grupo implica frecuentemente el empleo de técnicas como el seguimiento de mirada, o de conceptos como la saliencia, así como el uso de técnicas de aprendizaje profundo.

1.3. Planificación y herramientas

Para la elaboración de este trabajo de fin de grado, se ha realizado una división de tareas en función de la estructura del proyecto.

En la *fase de documentación*, se estudia el estado del arte referido a la conducción, los tipos de ADAS empleados, las técnicas que estos utilizan, y aplicaciones del seguimiento de la mirada

en ADAS. Para ello, se realiza una amplia revisión bibliográfica de libros, artículos, normativa, y contenido web, que se encuentra mayoritariamente reflejada en los Capítulos 2 y 3 de esta memoria.

La *fase de desarrollo inicial* consiste en la introducción inicial al aprendizaje profundo mediante la consulta de diversas fuentes bibliográficas que guardan relación con el proyecto [5, 6, 7], relacionadas con las herramientas a utilizar [8, 9], con la intención de impulsar el desarrollo de las habilidades requeridas en posteriores etapas del proyecto.

La *fase de desarrollo* engloba la parte de aplicación práctica del proyecto, en la cual se crea el modelo empleado, diseñando su arquitectura original sobre la cual se desarrollan diferentes versiones a través de la posterior implementación de diversas mejoras. Para estas tareas se ha empleado el lenguaje de programación *Python* junto con la librería de aprendizaje profundo *PyTorch* [10] para el sistema de precicción de la atención, y la librería *Albumentations* [11] para el aumento de datos. El código del proyecto está correctamente estructurado en un repositorio de GitHub¹. El trabajo se ha desarrollado en un ordenador con procesador Intel i7-8750H, 16GB de RAM y tarjeta gráfica Nvidia GeForce GTX 1050.

La *fase de redacción* está compuesta por las etapas de escritura de los distintos capítulos de la memoria. La composición de la misma se ha realizado de forma escalonada y progresiva, conforme se ha avanzado en las fases anteriormente explicadas. Para la elaboración de esta memoria, se ha utilizado el editor web *Overleaf*, que emplea el procesador de texto *LaTeX*, usado para la escritura y edición de documentos científicos.

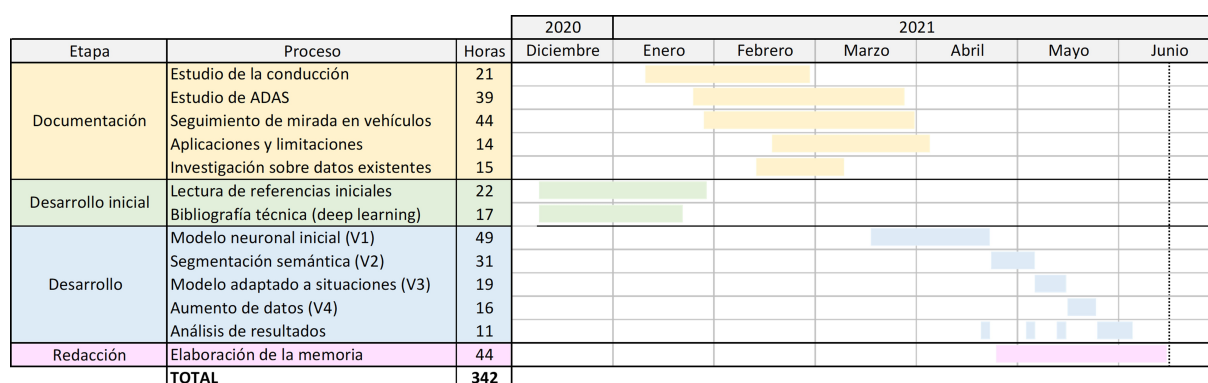


Figura 1.2: Diagrama Gantt con las horas dedicadas a cada tarea del proyecto. La línea discontinua indica la semana de entrega de la memoria.

Este proyecto se ha desarrollado de forma iterativa (ver Figura 1.2), contando con un seguimiento constante a través de reuniones de control periódicas. En estas reuniones, se mostraba el progreso realizado en las distintas fases, se planteaban cuestiones, se realizaban discusiones, y finalmente, se proponía una planificación con los siguientes pasos a seguir.

¹<https://github.com/BlancaLH/TFGCodigo>

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

En la actualidad, la mayoría de los vehículos que se encuentran en circulación han sido equipados con Sistemas Avanzados de Asistencia al Conductor (ADAS), sistemas diseñados con la finalidad de apoyar la conducción del vehículo mediante la advertencia al conductor, la cooperación con el mismo, o incluso la toma de control del vehículo ante situaciones de riesgo o cambios en el entorno [12].

Uno de los objetivos de este proyecto es revisar el estado del arte en ADAS, y así proporcionar una visión completa, estructurada y coherente de las principales características que definen a los ADAS actuales, que posteriormente permita analizar potenciales mejoras a los mismos. En este capítulo, se revisa el concepto de autonomía del vehículo en función de las características del mismo, así como los distintos niveles de autonomía posibles. Seguidamente, se estudia el papel de los ADAS que se implementan actualmente en vehículos, proponiendo una taxonomía que permita clasificar los distintos tipos de ADAS actuales, y analizando tanto los sensores que forman parte de ellos como los flujos de datos y sistemas de alerta y respuesta a ellos asociados.

2.1. Vehículo autónomo

La conducción autónoma es la imitación y reproducción, mediante un Sistema de Conducción Autónoma (*Automated Driving System*, ADS), de las capacidades humanas de manejo y control de un vehículo. Por tanto, se denomina vehículo autónomo a aquel que está dotado de un ADS, compuesto por el *hardware* y *software* necesario para llevar a cabo todas las funciones relacionadas con la tarea de conducción dinámica (es decir, en tiempo real) sin necesidad de intervención del conductor una vez haya fijado el destino.

2.1.1. Niveles de autonomía (*SAE*)

Los vehículos pueden clasificarse según su nivel de autonomía, dependiendo del grado de asistencia que ofrecen al conductor. Para ello, la Sociedad de Ingenieros de la Automoción (en inglés, *Society of Automotive Engineers*, *SAE*¹), o SAE Internacional, establece los aspectos que han de tenerse en cuenta tanto en el proceso de fabricación como en la regulación de este

¹<http://https://www.sae.org/>

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

Nivel	Denominación SAE	Ejecución de la dirección, aceleración y deceleración	Monitorización del entorno	Ejecución alternativa de la conducción	Modos de conducción disponibles en el sistema
0	Automatización nula	Conductor	Conductor	Conductor	N/A
1	Asistencia al conductor	Conductor y sistema	Conductor	Conductor	Algunos modos de conducción
2	Automatización parcial	Sistema	Conductor	Conductor	Algunos modos de conducción
3	Automatización condicional	Sistema	Sistema	Conductor	Algunos modos de conducción
4	Alta automatización	Sistema	Sistema	Sistema	Algunos modos de conducción
5	Automatización total	Sistema	Sistema	Sistema	Todos los modos de conducción

Tabla 2.1: Niveles de autonomía de la conducción, SAE Internacional [13]. La línea negra indica la frontera en la cual pasa a ser el sistema el que realiza ciertas tareas.

tipo de vehículos, y plantea una clasificación en función del nivel de autonomía de los mismos [13].

Existen seis niveles de autonomía del vehículo definidos originalmente por la SAE en el año 2014, que han sido revisados y actualizados en años posteriores [14, 15], y numerados en orden creciente de menor a mayor autonomía, siendo el nivel 0 aquel sin automatización de la conducción, y el nivel 5, aquel que proporciona una automatización completa.

Como se observa en la Tabla 2.1, los niveles 0 al 2 corresponden a vehículos en los que el conductor analiza el entorno de la conducción, y los niveles 3 al 5 implican que el ADS se encarga de la tarea de monitorización del entorno.

El nivel 0, o de *automatización nula*, implica que el conductor efectúa en todo momento el manejo del vehículo y realiza la conducción dinámica (dirección, aceleración y frenado) del vehículo. En todo caso, los ADAS podrían alertar al conductor, pero no se automatiza ninguna tarea de la conducción de forma activa.

El nivel 1, o de *asistencia al conductor*, requiere una monitorización constante del entorno por parte del conductor, ya que es quien toma la decisión de activar o desactivar el ADAS. En este nivel, los sistemas permiten controlar la tarea de conducción dinámica longitudinal, mediante la regulación de la velocidad, o lateral, a través del ajuste de la dirección, siempre y cuando hayan sido activados por el conductor. El control que realiza el sistema en este nivel es parcial, no pudiendo controlar parámetros laterales y longitudinales de forma coordinada en un mismo instante. La desactivación de estos sistemas se produce en el instante en el que el conductor decide tomar el control del vehículo.

El nivel 2, o de *automatización parcial*, permite que los ADAS controlen la conducción dinámica longitudinal y lateral de forma simultánea en caso de haber sido activada. Es imprescindible que el conductor monitorice el entorno de la conducción constantemente, pudiendo activar y desactivar el ADAS en cualquier momento.

El nivel 3, o de *automatización condicional*, se caracteriza porque un ADS controla totalmente la conducción dinámica del vehículo, encargándose de la monitorización del entorno, sin esperar intervención del conductor, aunque sí una atención activa hacia el entorno y el sistema, como medida de seguridad, y con la finalidad de poder responder en el menor tiempo posible en caso de ser alertado. En caso de falta de atención, el sistema podría tomar el control del vehículo en condición de operación de riesgo mínimo, implicando el manejo del vehículo hasta su detención

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

SAE	Automatización nula (Nivel 0)	Asistencia al conductor (Nivel 1)	Automatización parcial (Nivel 2)	Automatización condicional (Nivel 3)	Automatización elevada (Nivel 4)	Automatización total (Nivel 5)
BASt	Sólo el conductor	Asistido	Parcialmente automatizado	Altamente automatizado	Totalmente automatizado	N/A
NHTSA	Sin automatización (Nivel 0)	Automatización de funciones específicas (Nivel 1)	Automatización de funciones combinadas (Nivel 2)	Automatización para la conducción autónoma limitada (Nivel 3)	Automatización para la conducción autónoma total (Nivel 4)	

Tabla 2.2: Correspondencia entre los niveles de automatización SAE, BASt y NHTSA [13].

segura de tal forma que el riesgo de accidente fuese minimizado. No obstante, esta función de emergencia puede verse limitada por factores como el tipo de sistema empleado o el entorno. Además, el ADS permite únicamente su activación bajo las condiciones para las cuales ha sido diseñado, y es el conductor quien ha de determinar si su activación es apropiada.

El nivel 4, o de *alta automatización*, emplea varios ADS que controlan todos los aspectos de la conducción dinámica, incluso en caso de que el conductor no reaccione ante una petición de intervención por parte del sistema. Estos sistemas están restringidos a zonas geográficas en las que existan bases de datos con información del entorno, como ciertas carreteras. En este caso, el conductor no necesita monitorizar activamente el entorno o las acciones del ADS, pero ha de estar preparado para asumir el control del vehículo en el momento que el sistema lo requiera. En este último caso, el conductor dispone de un tiempo de respuesta del orden de segundos, tras el cual deberá retomar activamente la conducción dinámica. En caso de que no haya una respuesta por parte del conductor, el vehículo responde entrando en condición de operación de riesgo mínimo, controlando el vehículo en todo momento.

Finalmente, el nivel 5, o de *automatización total*, se fundamenta en un ADS que realiza la tarea de conducción dinámica de forma totalmente automatizada independientemente de las condiciones del entorno. El ADS puede llevar a cabo cualquier tarea propia del conductor, incluyendo la reacción ante emergencias o fallos de sistema que impliquen entrar en condiciones de operación de riesgo mínimo. En este caso, el conductor puede ser prescindible, aunque también puede activar y desactivar el sistema manualmente.

2.1.2. Otras clasificaciones de niveles de autonomía

Es importante destacar que, además de la clasificación anterior, propuesta por la SAE, existen otras estructuraciones de los niveles de autonomía del vehículo, como la descrita por la Administración Nacional de Seguridad del Tráfico en las Carreteras (NHTSA) en Estados Unidos, o la elaborada por el Instituto Federal Alemán de Investigación de Carreteras (BASt).

Si bien los niveles de autonomía descritos son consistentes entre los tres documentos, cabe mencionar que no han sido descritos de manera idéntica (ver Tabla 2.2). Es por ello que en la clasificación BASt se omite el nivel 5, presentando así un nivel menos, y se modifica la descripción del nivel 4, que hace referencia al vehículo totalmente automatizado. Por tanto, el nivel BASt 4 equivale al nivel SAE 5 de automatización total. Del mismo modo, la principal diferencia entre la clasificación de la SAE y la de NHTSA es que esta última agrupa en el nivel 4 los niveles SAE 4 y 5. Además, las definiciones de la NHTSA para cada nivel están adaptadas a

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

la normativa estadounidense para vehículos autónomos, siendo esta todavía inexistente como normativa específica en Europa.

2.1.3. Nivel de autonomía actual

A la hora de determinar el nivel de autonomía en el que se encuentran los vehículos, se plantean hipótesis en las que los niveles se espacian 15 años entre sí, como las formuladas a partir de datos experimentales por Rödel et al. [16], que recorren los niveles de autonomía según la NHTSA desde el año 1985 hasta 2045. Originalmente, los vehículos no presentaban ningún tipo de automatización. En el año 2000, se comienzan a implementar los sistemas de navegación y la transmisión automática. El control de cruce adaptativo y el asistente de estacionamiento se integran en 2015, alcanzando así el nivel 2 de automatización. Por otro lado, se prevé que el vehículo conduzca de forma autónoma en autopistas para el escenario de 2030, llegando al nivel 3 de automatización. Y, finalmente, se plantea la potencial existencia de un vehículo completamente autónomo en 2045.

Por todo ello, es comúnmente aceptado que el grado de automatización actual de los vehículos corresponde a los niveles SAE 2 y 3 [17, 18, 19, 20], con sistemas como el *DrivePilot* de Mercedes [21] o el *ProPilot* de Nissan [22] característicos del nivel 2, o con el célebre sistema AutoPilot 2.0 [23] considerado como avance hasta el nivel 3, al mismo tiempo que se produce un desarrollo de prototipos que alcanzan el nivel SAE 4 [24]. Sin embargo, dada la variabilidad entre distintas escalas, así como el desarrollo de tecnologías que cumplen solo parcialmente los requisitos establecidos por las mismas en su fase de desarrollo inicial, no existe una única atribución en cuanto a nivel de automatización completamente correcta.

El desarrollo de ADAS en estos niveles de autonomía es progresivo, y se nutre de sistemas que satisfacen parcialmente los criterios propios de los niveles, ya que han de garantizar la seguridad vial e intrínseca del vehículo, entre otros factores. Por ello, a día de hoy se pueden identificar todavía diversas carencias, como la necesidad de bases de datos con grandes volúmenes de información del entorno (que limita el área de actuación y la precisión del ADAS), como ocurre con Tesla [25, 26]; o de carreteras compatibles con el sistema, como es el caso del sistema *Super Cruise* de Cadillac [27].

2.2. ADAS en vehículos

Tal y como se ha analizado en la Sección 2.1, actualmente, la incorporación de ADAS en vehículos facilita notablemente el desempeño de la conducción, teniendo un papel fundamental en la automatización de ciertas acciones del vehículo. Para ello, los ADAS monitorizan ciertas variables del entorno o del conductor, procesan la información capturada, y, finalmente, generan respuestas con mayor o menor autonomía, que pueden incluir desde señales luminosas y acústicas en el interior del vehículo en los niveles 1 y 2, hasta la toma de control de parámetros de conducción (como la dirección o la velocidad) por parte del sistema en los niveles 2 y 3.

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

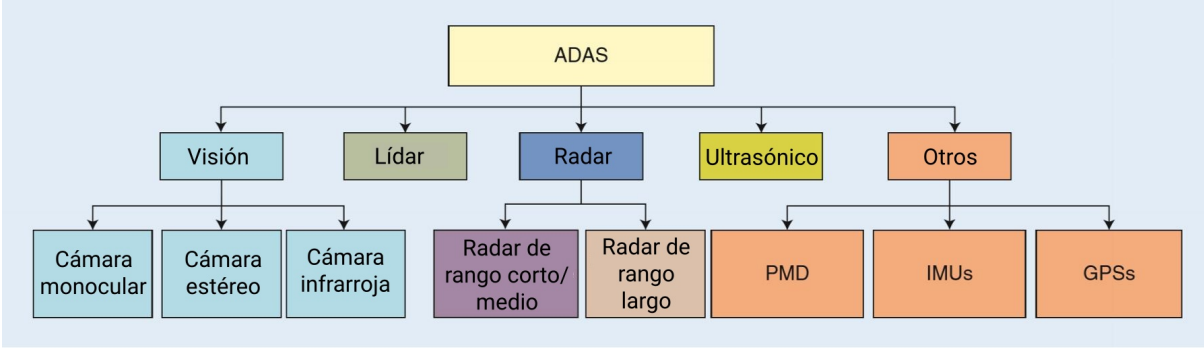


Figura 2.1: Taxonomía de un ADAS en función de los tipos de sensores que emplea [4].

2.2.1. Tipos de ADAS

Existen numerosos métodos de clasificación de ADAS en función de diversos criterios. Entre ellos, se encuentran el nivel de automatización, el aspecto económico, o el tipo de respuesta que se genera, entre otros. En este proyecto, se estudia la taxonomía propuesta por Kukkala et al.[4], basada en la agrupación de ADAS en función de los sensores que utilizan.

A la hora de realizar una clasificación de ADAS siguiendo dicho criterio, es de suma importancia tener en cuenta que, a su vez, los sensores pueden agruparse en *propioceptivos*, si detectan y responden ante situaciones de riesgo analizando parámetros del funcionamiento del vehículo (como la velocidad, aceleración o dirección); y *exteroceptivos*, cuando son capaces de responder con antelación mediante la predicción de posibles peligros al captar la información del exterior del vehículo. Existen también las *redes de sensores*, formadas por plataformas de multisensado que monitorizan el tráfico, y se encuentran en el entorno [28].

Tal y como se observa en la Figura 2.1, los sensores que emplea un ADAS actualmente están basados en visión (ver Subsección 2.2.2), lídar, radar, ultrasónicos, o en otras tecnologías. La ubicación de estos sensores en el vehículo se realiza de forma que el área en torno al vehículo quede cubierta, evitando la aparición de ángulos muertos o puntos ciegos. En la Figura 2.2 se observan los principales sensores empleados en ADAS implementados actualmente en vehículos.

2.2.2. Sensores

Los sensores de *visión* más empleados en ADAS actuales son las cámaras, que se encargan de realizar la captura de imágenes, obteniendo información sobre el color, el contraste, o la textura de los elementos que en ellas aparecen. Posteriormente, y mediante el empleo de sistemas embebidos, las imágenes son analizadas con múltiples finalidades, como detectar y reconocer elementos, segmentar el entorno o rastrear ciertas entidades. Además, debido a su bajo coste y fácil instalación, junto con el surgimiento de nuevas legislaciones, la integración de cámaras en vehículos modernos está experimentando un crecimiento desde el año 2018 [29]. En la actualidad, se emplean cámaras para monitorizar tanto el interior como el exterior en vehículos. Las cámaras más utilizadas en ADAS basados en visión son las monoculares, las estereoscópicas y las infrarrojas, cuyas características se detallan en la Subsección 3.1.1.

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

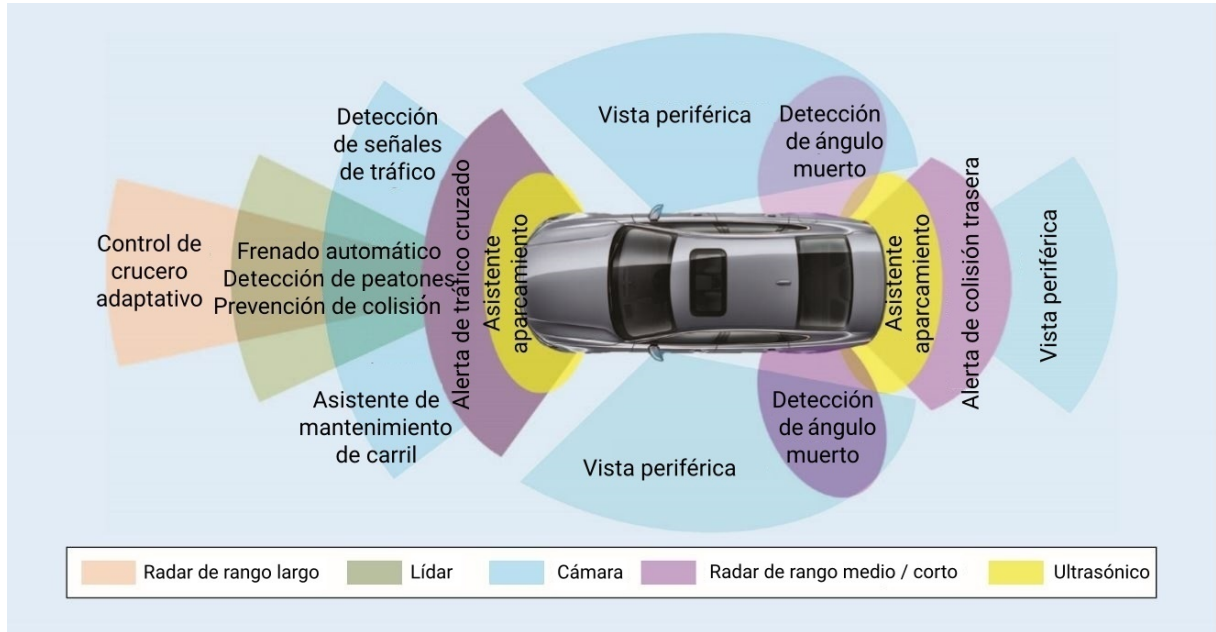


Figura 2.2: Disposición de sensores en el vehículo de tal forma que la mayor parte de su entorno quede cubierta [4].

El comúnmente denominado *líd*ar (*Laser Imaging Detection and Ranging, LIDAR*) es un dispositivo que emite un haz láser, y mide el tiempo que tarda en rebotar en un elemento del entorno y volver al sensor, con la finalidad de calcular la distancia entre ambos. Estos sistemas obtienen imágenes 3D de elevada resolución, y funcionan en rangos mayores que los sistemas basados en visión. Algunos *líd*ar incluyen sensores de vista periférica que emiten haces láser en todas las direcciones y de forma continuada. De este modo, se pueden generar imágenes 360°, tridimensionales, en las que se recoge el área circundante al vehículo con información de profundidad de gran precisión.

Los estudios realizados en prototipos [30, 31] han demostrado las principales ventajas de incorporar el *líd*ar a un vehículo, como la detección de objetos, la implementación del frenado automático, o la prevención de colisiones entre otras. Estos avances han provocado un aumento de popularidad del *líd*ar, siendo instalado en un gran número de vehículos en la actualidad.

A pesar de las características mencionadas anteriormente, los *líd*ar tienen ciertos inconvenientes, como su peso, tamaño y coste. Además, las condiciones atmosféricas como la lluvia o la niebla pueden comprometer la precisión y robustez de estos sistemas. A día de hoy, existen *líd*ar de estado sólido [32], considerablemente más ligeros y compactos, resultando a su vez más económicos, y facilitando así su inclusión en ADAS de vehículos de distintas gamas.

El *radar* es un sistema que emite microondas para estimar la velocidad y distancia a la que se encuentra un objeto a través de mediciones de variación de frecuencia en la onda que se refleja tras rebotar en este, así como por efecto Doppler. Al poseer mayor longitud de onda que la luz, empleada en *líd*ar, las microondas pueden cubrir rangos mayores de distancia, detectando así objetos más lejanos. A diferencia del *líd*ar, el rendimiento del radar no se ve afectado por condiciones atmosféricas como la lluvia o la niebla, y su coste es significativamente inferior. Sin embargo, el radar no cuenta con la sensibilidad que sí posee el *líd*ar, pudiendo realizar este

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

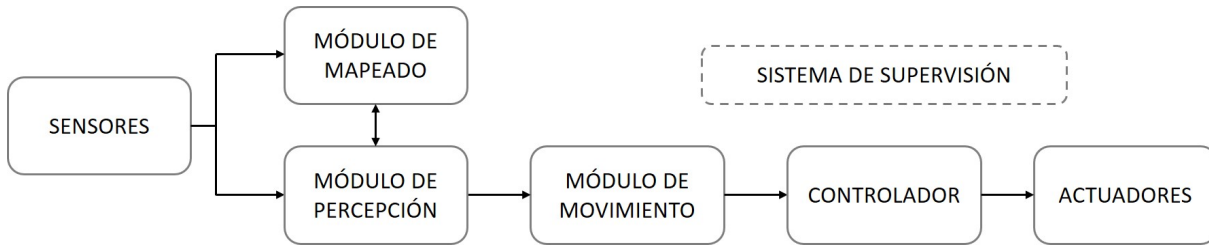


Figura 2.3: Esquema con los principales módulos que frecuentemente conforman un ADAS genérico, basado en el trabajo de Cova-Rocamora y Torres-Medina [35].

último una detección con mayor detalle en distancias largas. En función del área que cubren durante su funcionamiento, estos sistemas pueden ser de rango corto (0.2-30 m), medio (30-80 m), o largo (80-200 m) [33].

Las principales aplicaciones de radares de corto o medio alcance incluyen alertas de tráfico cruzado o detección del punto ciego, por lo que se encuentran ubicados en las esquinas del vehículo. Los radares con un rango mayor se emplean para el control de cruce adaptativo, siendo generalmente instalados bajo la calandra o rejilla delantera, o por debajo del parachoques del vehículo.

Los sensores *ultrasónicos* emplean ondas acústicas para medir la distancia a la que se encuentra un objeto. Se utilizan mayoritariamente para la detección de objetos muy próximos al vehículo, en aplicaciones como el asistente de aparcamiento automático o el asistente de estacionamiento en paralelo, y se integran bajo los parachoques trasero o delantero.

Por último, existen *otros sensores* que se emplean con la finalidad de mejorar las funcionalidades de los anteriores, tales como el *mezclador fotónico* (PMD), que analiza señales luminosas incoherentes, o la *unidad de medición inercial* (IMU) y el *sistema de posicionamiento global* (GPS), que contribuyen a una mejor medición de las distancias en combinación con el lidar o el radar.

2.2.3. Flujo de datos de un ADAS

El tratamiento de la información desde su captura a través de los sensores hasta la respuesta del sistema hacia el conductor mediante los actuadores varía en función de las características de cada ADAS. No obstante, existe una estructura general formada por cinco módulos básicos que suelen conformar estos sistemas (ver Figura 2.3) [34, 35].

El módulo de *percepción* es el encargado de evaluar el estado del vehículo en todo momento, así como de detectar los objetos del entorno, sean dinámicos o estáticos. El módulo de *mapeado* del entorno permite crear mapas en función de distintos criterios como la ocupación, la localización o las carreteras. El módulo de *planificación del movimiento* traza la ruta a seguir, siendo capaz de calcular y ofrecer diferentes alternativas en caso de contratiempo. El *controlador de parámetros* del vehículo obtiene información sobre los valores de aceleración, dirección, giro y frenado del vehículo. Por último, el *sistema de supervisión* revisa el correcto funcionamiento de los demás módulos, además de supervisar los sensores y actuadores.

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

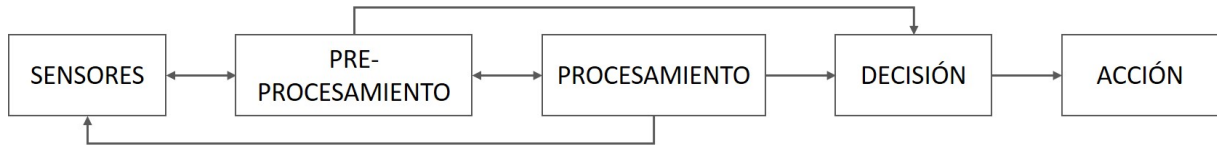


Figura 2.4: Diagrama general del flujo de datos de un ADAS desde su obtención hasta la respuesta del sistema.

Como se aprecia en la Figura 2.4, toda la información obtenida a través de los sensores es transmitida a la unidad de control del sistema, que procesa los datos y los contrasta con la información del vehículo, del conductor y del propio sistema, con la finalidad de identificar situaciones de riesgo y llevar a cabo una toma de decisiones. En caso de detectar algún tipo de riesgo, el sistema responde activando los actuadores pertinentes, comunicándose así con el conductor del vehículo [36].

Basados en los módulos y secuencias previamente comentados, y en función del grado de sofisticación del ADAS, existen múltiples arquitecturas, como sistemas de dos canales, en los que se establece una comparación entre dos unidades de control; o los sistemas 2oo3, que cuentan con tres unidades de control y un selector. Los sistemas 1oo2D de dos o más canales cuentan con grandes ventajas, ya que en caso de fallo operacional en uno de los canales, el sistema no se apaga y continua funcionando con un único canal [37].

Adicionalmente, la arquitectura tiene una gran repercusión en la conducción en caso de fallo del sistema si el ADAS entra en condición de operación de mínimo riesgo: Según características como su robustez, o el número de canales y unidades de control del sistema, el ADAS podría desconectarse y apagarse por completo o, por el contrario, seguir siendo funcional e intrínsecamente seguro, garantizando así una mayor seguridad del vehículo.

2.2.4. Sistemas de alerta/respuesta

La última fase de la tarea que realiza un ADAS se basa en la interacción persona-máquina (*Human-Machine Interface, HMI*) [38], ya que el sistema se encarga de notificar al conductor cualquier evento relevante que pueda afectar a la tarea de conducción, o incluso entrañar un riesgo. En este caso, la actuación del sistema puede ser en forma de alertas al conductor o, en los casos más avanzados, mediante la toma de control del vehículo de forma temporal.

Las notificaciones y alertas al conductor suelen darse en forma de señales acústicas o luminosas en pantalla, luna delantera o retrovisores, tanto en el interior como exterior del vehículo, como ocurre en los sistemas de alerta de tráfico cruzado en la parte trasera del vehículo (*Rear Cross Traffic Alert, RCTA*) [39]. Algunas versiones más sofisticadas de ADAS pueden llegar a controlar el frenado del vehículo al mismo tiempo que se realiza una alerta sonora [40].

Otros ADAS más avanzados realizan alertas más complejas en función de la respuesta del conductor [41, 42]. El anteriormente mencionado sistema *Super Cruise* de Cadillac (ver Subsección 2.1.3) posee tres niveles de alerta: En el primero se produce un aviso luminoso en el volante; en el segundo se emite una señal acústica y se activa la vibración del asiento del conductor, requiriendo una respuesta activa en el volante por parte del conductor; mientras que en el tercero se realiza una petición por voz de accionamiento manual del volante. En caso de no obtener respuesta por parte del conductor, el vehículo disminuye su velocidad hasta detenerse

2. Sistemas Avanzados de Asistencia al Conductor (ADAS)

en un lugar seguro, donde llamará a un asistente [27].

La incorporación de ADAS que alertan al conductor representa un punto de inflexión en el desarrollo de sistemas para la automatización de la conducción, puesto que suponen un apoyo a las tareas del conductor, reduciendo así los riesgos que conllevan. Por tanto, es importante conocer el impacto que representan estas alertas en el conductor, así como su interacción con el ADAS.

Las reacciones de un conductor ante diferentes ADAS pueden variar notablemente, demostrándose que el empleo de estos sistemas puede tener un efecto negativo en caso de que el conductor no haya interiorizado un esquema de funcionamiento del ADAS [43]. Este tipo de respuestas se consideran limitaciones del sistema, aunque no lo supongan a nivel técnico. Por ello, es de suma importancia que el conductor conozca e interiorice el funcionamiento de los ADAS del vehículo que conduce, así como su relación con los mismos.

A raíz de estudios de monitorización del estado del conductor [44], se ha demostrado que el empleo de técnicas poco o no invasivas facilita la detección del estado mental del conductor, que contribuye a su vez en la mejora del tipo de alertas que ha de producir un ADAS para obtener la reacción esperada en el conductor, en función de cada situación.

3. ADAS basados en visión

Como se ha visto en el capítulo anterior, los ADAS basados en visión son aquellos que dependen de imágenes obtenidas mediante cámaras, de las cuales extraen información mediante procedimientos de visión por computador. Estos sistemas procesan, analizan e interpretan las imágenes de forma automática.

En este capítulo, se profundiza en ADAS basados en *visión*, estrechamente relacionados con el objeto de este proyecto, y siguiendo una taxonomía basada en la escena que monitorizan, distinguiendo entre aquellos que controlan los eventos del interior o del exterior del vehículo. Además, se explican los principales sensores que emplean, y se realiza un análisis del procesamiento de información en este tipo de sistemas. Finalmente, se detallan varios casos prácticos que ejemplifican los conceptos mencionados a lo largo del capítulo.

3.1. Captura de información

En el capítulo anterior se ha estudiado cómo la captura de la información en ADAS se realiza mediante sensores. Específicamente, en el caso de ADAS basados en *visión*, se emplean cámaras, que se encargan de la monitorización de la escena, ya sea exterior o interior, en todo momento. Las cámaras pueden ser de diversos tipos (monoculares, estereoscópicas o infrarrojas), en función de la finalidad o ámbito de funcionamiento del sistema.

3.1.1. Sensores

La versatilidad, disponibilidad y precio moderado de las cámaras son los principales factores que influyen en que sean los sensores más empleados en este tipo de sistemas. En función del presupuesto, la gama del vehículo en el que se implementa, y la función del ADAS, se recurre generalmente a la instalación de uno de los tres tipos de cámara previamente mencionados.

Las *cámaras monoculares* se caracterizan por estar formadas por una única lente. Los sistemas que emplean este tipo de cámaras solo son capaces de obtener una imagen en cada instante temporal, por lo que los requerimientos para su procesamiento son bajos. Las cámaras monoculares pueden ser usadas para múltiples aplicaciones, como la detección de obstáculos, peatones, carriles y señales de tráfico [45]. Además, pueden emplearse para monitorizar al conductor en el interior de un vehículo, por ejemplo, para analizar la posición de su cabeza mediante la detección de la cara y los ojos (ver Figura 3.1b) [46]. Sin embargo, las imágenes monoculares son



(a) Cámara estereoscópica tras el retrovisor central, orientada hacia el frente en un vehículo [49].

(b) Cámara interior orientada al conductor [50].

Figura 3.1: Disposición usual de las cámaras en el vehículo.

capturadas desde un único punto de vista. Esto impide cualquier proceso de triangulación o estimación de profundidad de los elementos de la escena, por lo que este tipo de sensores no puede ser empleado para ninguna funcionalidad que requiera estimar distancias.

Las *cámaras estereoscópicas* están compuestas por dos o más lentes, con sensores de imagen en cada una de ellas, separadas una cierta distancia conocida como *base estereoscópica*. Al tomar múltiples perspectivas bidimensionales (2D) [47] de una misma escena tridimensional (3D), es posible emplear técnicas de visión por computador (como *geometría epipolar* o *structure-from-motion* (sfm)) para obtener información 3D (por ejemplo, profundidad o distancias) de la escena. Las cámaras estereoscópicas pueden tener aplicaciones análogas a las monoculares, aunque ofreciendo una mayor precisión; además de posibilitar otras como la estimación de distancias de los elementos del entorno. En la Figura 3.1a se observa su disposición a la altura del espejo retrovisor central, ya que en esa zona se maximiza la luz captada sin generar interferencias en el rango de visión del conductor [48], capturando la vista frontal del vehículo hacia la vía.

Las *cámaras infrarrojas* (IR) reciben su nombre por usar luz infrarroja, un tipo de radiación fuera del espectro visible por el ser humano. Se pueden dividir en dos subtipos: activas y pasivas. Las cámaras IR activas usan una fuente cercana de luz infrarroja incorporada en el vehículo para iluminar la escena, y un sensor de cámara digital para capturar la luz reflejada por los elementos de la escena. Por otra parte, las cámaras IR pasivas están equipadas con un sensor IR en el que cada píxel puede ser considerado como un sensor de temperatura que captura la radiación térmica emitida por cualquier material. A diferencia de las cámaras IR activas, las cámaras IR pasivas no necesitan ningún tipo de iluminación especial de la escena. A pesar de ello, las cámaras IR activas son todavía las más empleadas para la asistencia al conductor en condiciones nocturnas, ya que permiten capturar información de la escena incluso en condiciones de iluminación escasa. Este tipo de cámaras puede emplearse en la monitorización del tráfico [51], de vehículos, peatones u otros elementos en condiciones de menor visibilidad.

3.1.2. Monitorización

La monitorización consiste en la observación y control del desarrollo de una acción a través del sistema. Concretamente, el ADAS toma la información de la escena a través de los sensores



Figura 3.2: Procedimiento de reconocimiento de señales mediante la localización de la corona circular roja exterior, obtenida por segmentación [53].

y la procesa, con la finalidad de detectar posibles anomalías en la misma. En función de la escena que se captura, la monitorización puede darse tanto en el exterior como en el interior del vehículo [4, 52].

Monitorización exterior

La monitorización exterior (*outdoor monitoring*) consiste en la captura de imágenes del entorno del vehículo para clasificar todo elemento físico externo al mismo, como peatones, vehículos y carreteras. Estos sistemas se distribuyen de tal forma que se cubra la mayor área posible del vehículo, haciendo especial hincapié en las zonas que se encuentran más comprometidas (frontal y trasera), así como en los puntos ciegos del vehículo. En muchas ocasiones, estos sistemas basados en visión se complementan con otros (lídár, radar), que prestan apoyo en condiciones desfavorables, aportando una mayor fiabilidad y garantía de seguridad al sistema.

La *detección de peatones* se basa en métodos de clasificación en función de su figura de forma simultánea, ya que la configuración en la que aparecen en la escena es variable, pudiendo no ser reconocidos por un único sistema en todo momento [54]. Con el avance de las técnicas de aprendizaje automático, se han desarrollado sistemas basados en redes neuronales convolucionales (ver Anexo A) capaces de identificar peatones, e incluso clasificar sus acciones [55].

La *detección de vehículos* es uno de los principales enfoques de la identificación de objetos en ADAS. Es de gran relevancia que los vehículos compartan características comunes entre sí (ruedas, luces de freno, matrículas), ya que resulta ventajoso a la hora de detectar la presencia de un vehículo a partir de la identificación parcial del mismo. De hecho, son estas particularidades que comparten los vehículos las que permiten su diferenciación respecto a otros elementos de la escena, como señales de tráfico o carreteras. La orientación de los vehículos puede generar dificultades a la hora de su identificación, ya que un vehículo visto de frente presenta un conjunto de características distinto al de un vehículo captado lateralmente. Por ello, los sistemas de clasificación de vehículos recientemente basados en redes neuronales convolucionales junto con sistemas de detección en tiempo real [56] son capaces de considerar diferentes tipos de vehículos desde perspectivas distintas.

Durante los últimos años, el *reconocimiento de señales de tráfico* ha ganado popularidad entre los ADAS. El caso más frecuente es la detección de límites de velocidad mediante la lectura de señales de tráfico que indican los límites de velocidad en la carretera. En caso de ser sobrepasados, el ADAS alerta al conductor. En la mayoría de los casos, se identifican los bordes coloreados para ubicar las señales en la imagen (ver Figura 3.2), y se emplea reconocimiento de caracteres para la lectura del contenido de la misma [53].



Figura 3.3: Detección de bordes de imagen mediante algoritmo de Canny e identificación de los límites de carril con la transformada de Hough [59].

La *detección de carril* permite mantener el vehículo entre las líneas de carril en una carretera. Se trata de una característica de ADAS que todavía no ha sido implementada mayoritariamente en el mercado automovilístico. Esto se debe a que presenta grandes limitaciones en la detección de carril debido a importantes inconsistencias que se encuentran frecuentemente en muchas carreteras, como el color de la pintura, el mal estado o incluso la falta de líneas de carril. Como se observa en la Figura 3.3, la mayoría de métodos de detección de carril emplean algoritmos para localizar bordes en una imagen (por ejemplo, el *algoritmo de Canny* [57]), junto con técnicas para determinar cuáles corresponden al límite del carril (como la *transformada de Hough* [58]).

Otra de las aplicaciones de la monitorización externa es la evasión de colisiones mediante ADAS que incorporan la característica del frenado automático. Estos sistemas combinan métodos de seguimiento de objetos, detección de vehículos y cálculo de distancias para evitar colisiones frenando o modificando la dirección del vehículo (ver Figura 3.4).

Monitorización interior

Por otro lado, la monitorización interior (*indoor monitoring*) se centra en capturar información del conductor. Al mismo tiempo que los ADAS han ido adquiriendo mayor relevancia y presencia en los vehículos, la monitorización del conductor mediante el empleo de cámaras también ha ido en aumento. Esto se debe a que, según estudios como el realizado por la NHTSA [60], el 80 % de los accidentes de tránsito con vehículos son causados por fatiga, somnolencia o distracción del conductor. Por ello, se justifica la importancia de estos sistemas, que evalúan el estado de atención del conductor en todo momento y, en caso de que este no preste atención a la carretera (ya sea por distracción con el móvil, somnolencia, distracciones en el interior del vehículo, o causas médicas), alertan al conductor, e incluso intentan salir de la carretera para detenerse en una zona segura [46]. Otros sistemas de detección de fatiga y somnolencia (*Driver's Drowsiness Detection*) alertan mediante la emisión de vibraciones en el asiento del conductor junto con señales acústicas en el interior del vehículo [61].

3.2. Flujo de datos

Generalmente, en los ADAS basados en visión, se produce un flujo de datos divisible en seis etapas principales: adquisición de imágenes, preprocesado, segmentación, detección y seguimien-



Figura 3.4: Detección y seguimiento de objetos y vehículos, y cálculo de distancias como parte del sistema de evasión de colisiones [4].

to de objetos, estimación de profundidad y control del sistema.

El proceso de *adquisición de imágenes* consiste en la captura de fotogramas de una escena. En la mayoría de los casos, el fotograma es una matriz formada por los píxeles de la imagen, que posee tres canales de información: rojo, verde y azul (RGB). Una gran parte de los ADAS poseen frecuencias de muestreo que varían entre 5 y 60 fotogramas por segundo (fps). Esto se debe a que en aplicaciones como la detección de vehículos próximos, cuyo posicionamiento puede variar en un período corto de tiempo, se requieren frecuencias más altas; mientras que en otras como la detección de señales de tráfico, evento invariante en tiempos cortos, se precisa un único fotograma para la lectura de la señal, por lo que la frecuencia de muestreo puede ser menor.

El *preprocesado* consiste en el conjunto de tratamientos previos que se aplican sobre una imagen para, posteriormente, ser empleada en un algoritmo de visión por computador. Algunas de estas técnicas son la eliminación de ruido, la estabilización de imagen, la corrección de color o la normalización. Asimismo, en muchos casos, se realizan separaciones en luces y sombras, o zonas sobreexpuestas y subexpuestas, para facilitar las tareas de seguimiento y detección.

La *segmentación* es el análisis de una imagen para separar regiones de características similares en la misma. Para ello, se emplean técnicas de filtrado de imagen, que permiten distinguir los píxeles pertenecientes a un mismo elemento, como por ejemplo, la carretera o el cielo. Algunos de estos métodos emplean información de los colores para segmentar distintos elementos o patrones en los objetos (como la detección de señales de tráfico empleando algoritmos que filtran el color rojo, característico en bordes de las señales de tráfico de gran cantidad de países en Europa así como en Estados Unidos). El resultado de estas técnicas son imágenes binarias en las que todos los píxeles que no coinciden con los parámetros del filtro se vuelven negros, y los píxeles de interés, blancos, funcionando así como una máscara que permite encontrar el área de interés en la imagen original (ver Figura 3.2).

El proceso de *detección y seguimiento de objetos* en una imagen consiste en la clasificación y predicción del movimiento de los mismos. Para ello, se emplean algoritmos de *machine learning*, que requieren extensas bases de datos para aprender a identificar los distintos elementos de la imagen. Estas técnicas permiten desarrollar modelos capaces de extraer información latente

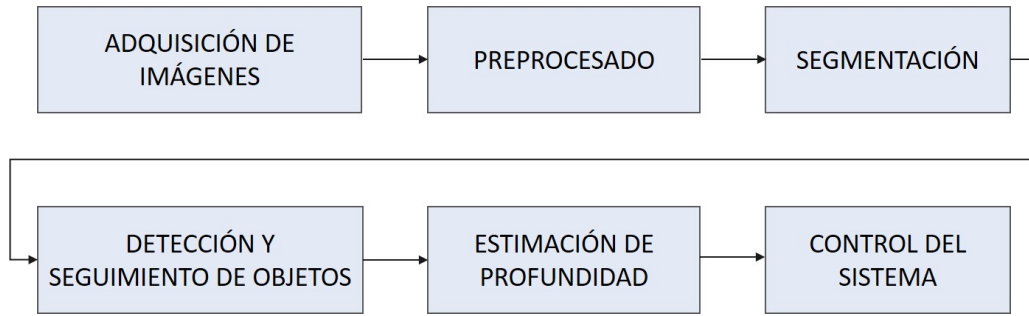


Figura 3.5: Flujo de datos con seis etapas en ADAS basados en visión [4].

de las imágenes preprocesadas por el ADAS, para clasificar los objetos e incluso predecir su comportamiento en base al conocimiento previo.

La *estimación de profundidad* consiste en el cálculo de la distancia entre un objeto y el sensor (la cámara) a partir de la imagen captada por la misma [62]. Para ello, se emplea una cámara estereoscópica, se estima un mapa 3D de profundidad y se realiza una reconstrucción de la escena, tal y como se ha mencionado en la Subsección 3.1.1.

El *control del sistema* es el último paso que recorren los datos en un ADAS, y es en el cual se interpretan los resultados obtenidos en las etapas previas. Para ello, se otorga un peso a cada etapa en función de su relevancia, lo que permite realizar una toma de decisiones ponderada con todos los parámetros que influyen en la misma, en función de las características del ADAS.

3.3. Respuesta del vehículo

Existe una gran variedad ADAS que emplean *visión por computador* para la monitorización del entorno del vehículo. Estos sistemas se distinguen por su polivalencia, pudiendo tener diversidad de aplicaciones al estar diseñados para cubrir las distintas necesidades del conductor durante el manejo el vehículo. Como se ha explicado en la Sección 2.2, la comunicación del sistema con el conductor es la última etapa en la que interviene un ADAS, para lo que se emplean los actuadores. Estos pueden avisar al conductor mediante notificaciones, alertas o la toma de control del vehículo.

La gran mayoría de sistemas realizan advertencias en forma de señales sonoras o luminosas. Un ejemplo a considerar es el RCTA¹, que emplea iluminación LED y alertas sonoras que se activan en situaciones de tráfico en maniobras complejas como el cambio de carril o el aparcamiento con visibilidad reducida (ver Subsección 2.2.4). Habitualmente, el RCTA se emplea en combinación con el sistema de detección de punto ciego (*Blind Spot Monitor*), que capta vehículos, obstáculos y peatones en la parte trasera y el lateral del vehículo, realizando una alerta visual y auditiva al mismo tiempo que emite vibración en el asiento del conductor [63].

Otros sistemas trabajan conjuntamente para realizar acciones más complejas, en los que la toma de decisiones se ve condicionada por diferentes factores. Un ejemplo es el sistema de monitorización del conductor o sistema de monitorización de la atención (*Driver Attention Mo-*

¹Sistema de alerta de tráfico cruzado (*Rear Cross Traffic Alert*), mencionado en el Capítulo 2.

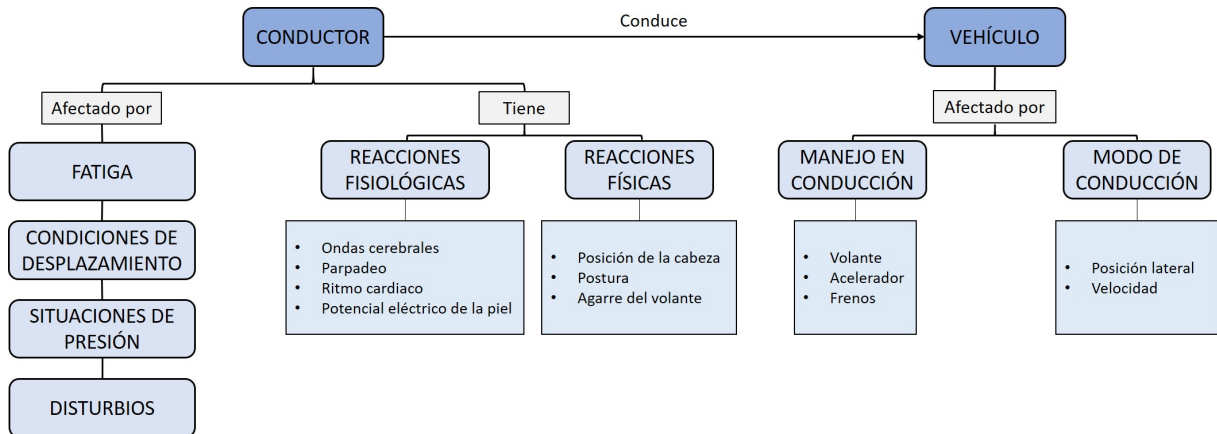


Figura 3.6: Aspectos determinantes para evaluar el nivel de fatiga del conductor [61].

nitor), que trabaja generalmente junto con el sistema de precolisión. Este emplea una cámara IR acoplada en el volante para monitorizar la atención del conductor. En caso de detectar la falta de atención a la carretera y una situación peligrosa, el sistema realizará una alerta mediante sonidos y luces intermitentes. En caso de que el conductor no responda, el vehículo tomará el control de los frenos.

Como se puede observar, algunos ADAS pueden llegar a tomar el control parcial o total de los parámetros de la conducción (aceleración, frenado, dirección) (ver Tabla 2.1). Es el caso del conductor asistente de emergencia (*Emergency Driver Assistant, EDA*), empleado en caso de emergencia médica, donde si el sistema decide que el conductor no es capaz de conducir con seguridad, toma el control de los frenos y del volante hasta la parada completa del vehículo [52].

El sistema de detección de somnolencia del conductor (*Driver Drowsiness Detection*) es empleado en prevención de accidentes por adormecimiento del conductor. Este concepto surge de la estimación que realizan diversos estudios en la que en torno al 20 % de los accidentes en carretera guardan relación con conductores que presentaban signos de fatiga en ese momento [52]. Para ello, se combinan distintas tecnologías, con la finalidad de obtener una detección precisa del estado del conductor: se monitoriza el patrón de conducción (que se vuelve más brusco y abrupto en caso de presentar somnolencia) junto con el control de posición del vehículo en el carril mediante una cámara. También se analizan los aspectos psicofísicos del conductor; es decir, su comportamiento en función de estímulos cambiantes que lo condicionan. Para ello, se emplea una cámara que registra la cara y los ojos del conductor, así como sensores corporales para la medición de parámetros fisiológicos (ver Figura 3.6).

A pesar de que existe una gran variedad de ADAS actualmente implementados en el mercado del automóvil, muchos de ellos todavía presentan limitaciones ya que requieren grandes bases de datos para su correcto funcionamiento, o presentan incompatibilidades en distintos entornos al encontrarse limitaciones en las rutas [27], lo que reduce considerablemente el ámbito en el cual su actuación es segura. Además, la existencia de otros requisitos como el precalibrado de los sistemas de monitorización interna en función del conductor, su coste, o la respuesta de estos ante distintas condiciones de luminosidad, puede limitar su aplicabilidad.

Actualmente, la implementación de ADAS basados en visión se realiza únicamente en vehícu-

los de gama alta [27]. No obstante, se pretende que este tipo de ADAS sea más económico, y que pueda ser implementado en vehículos dirigidos a segmentos con mayor número de clientes.

Por otra parte, se ha demostrado que el empleo de este tipo de ADAS ha provocado una disminución del número de accidentes de tráfico [61]. Además, con el tiempo, las mejoras en la eficiencia de la interacción entre los usuarios y los distintos sistemas automáticos han influido en que el número de vehículos que incluyen estos sistemas haya ido en aumento. En el futuro, a corto plazo, se prevé que la mayoría de vehículos pertenecientes a los segmentos B y C incluyan ADAS basados en visión.

3.4. Seguimiento de la mirada en ADAS

La detección y seguimiento de la mirada (*eye tracking*) son un conjunto de técnicas que, basándose en características únicas del ojo humano, permiten a una serie de sensores (por ejemplo, cámaras estereoscópicas, monoculares o IR) obtener información sobre la posición y desplazamientos oculares de las personas, esto es, saber a dónde están mirando en cada momento. Este tipo de técnicas tiene un amplio abanico de aplicabilidad, en parte gracias a su potencial para el análisis del estado emocional, la ocupación mental, o incluso las necesidades de una persona únicamente a partir del comportamiento de su mirada [64].

La conducción resulta un contexto en el cual la aplicación de este tipo de técnicas puede ser beneficiosa, ya que los datos de la mirada de un conductor pueden ser empleados para valorar su estado de atención en tiempo real, en parte motivado por la existencia de patrones o comportamientos comunes en la conducción. Un claro ejemplo es la tendencia del conductor a mirar al frente del vehículo en situaciones con una gran demanda cognitiva, reduciendo así la frecuencia de mirada al velocímetro, espejos retrovisores o, en líneas generales, la visión periférica. Estos comportamientos pueden generar fenómenos como la ceguera por falta de atención, la pérdida de conciencia situacional, o situaciones conocidas como *looked-but-failed-to-see* [65], en las que el conductor mira pero no ve determinadas acciones o elementos del entorno, pudiéndose provocar un accidente.

En esta sección se incide en la importancia del conductor como elemento clave en la tarea de la conducción, evaluando su comportamiento en función del nivel de atención que presenta. Además, se explican distintas técnicas de seguimiento de la mirada así como los algoritmos que se emplean para procesar los datos obtenidos en el proceso de conducción. Finalmente, se discuten las limitaciones actuales de este tipo de sistemas.

3.4.1. Estado y comportamiento del conductor

Se pueden diferenciar tres elementos fundamentales que intervienen en el proceso de conducción: el conductor, el vehículo y el entorno de conducción. Primero, el conductor ha de comprender el entorno en el que conduce (señales de tráfico, carriles), comúnmente denominado conciencia situacional. Esta etapa es crucial en la conducción, puesto que se necesita una comprensión del entorno para la correcta realización de la tarea de la conducción, y se compone por la percepción de elementos del entorno, la comprensión del significado de los elementos

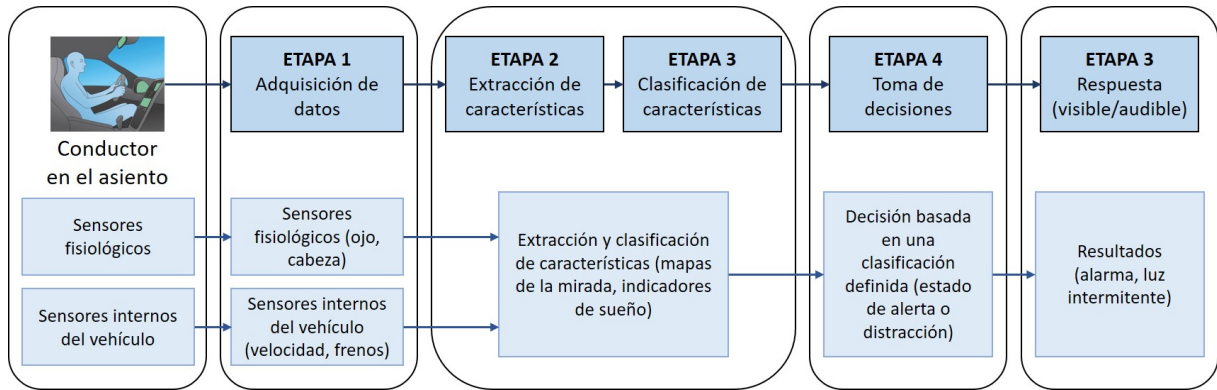


Figura 3.7: Etapas de procesamiento de datos de distinta naturaleza en algoritmos frecuentes en ADAS que monitorizan la mirada del conductor [66].

percibidos, y finalmente, la predicción de su impacto en un futuro próximo [66].

En segundo lugar, el conductor realiza una toma de decisiones, que puede involucrar múltiples eventos al mismo tiempo, para, finalmente, llevar a cabo una o varias acciones determinadas. Por todo ello, un ADAS que emplea *eye tracking* monitoriza continuamente la atención del conductor basándose en los comportamientos de su mirada, y genera una alarma o táctica de contrarresto en caso de negligencia (ver Figura 3.7).

Paralelamente, según estudios realizados en los últimos años [67], las principales causas de accidentes de tráfico son la distracción, la fatiga y la conducción agresiva, llegando a suponer un 90 % de los mismos. La fatiga implica una alteración en el rendimiento físico y mental que se manifiesta a través de la sensación de somnolencia en el conductor. Por otra parte, la conducción agresiva está ligada a acciones intencionadas del conductor, por lo que las señales de tráfico resultan más efectivas que la alerta proporcionada por un ADAS en este caso. La distracción, sin embargo, supone la implicación del conductor en una tarea ajena paralela a la conducción.

El comportamiento del conductor se ve profundamente afectado por distracciones, consideradas en muchos casos la razón de la mitad de los accidentes de tráfico [68, 69]. Dado el amplio rango de actividades que pueden desembocar en la distracción del conductor (comer, beber, utilizar el teléfono móvil, manipular elementos internos del vehículo, mirar el paisaje, etc.), la NHTSA realiza una categorización de las mismas en función del tipo de distracción que provocan en el conductor [70]: La distracción puede ser *visual* si implica la mirada, *cognitiva* si se refiere a la actividad mental, *física* si se trata de elementos corporales como las manos, o *auditiva* si se inhibe la captación de sonidos. Se puede conocer el nivel de atención del conductor mediante las acciones que este realiza durante la conducción, que actúan como indicadoras al reflejar su conducta. Las principales acciones que denotan el estado del conductor son el control lateral, el tiempo de reacción y la velocidad.

El *control lateral* del vehículo se ve afectado por la distracción visual del conductor, y se refleja en una mayor variabilidad del posicionamiento del vehículo en el carril, pues se producen mayores desviaciones en la trayectoria, compensados en muchas ocasiones mediante movimientos bruscos y repentinos del volante [71]. El *tiempo de reacción* se encuentra directamente relacionado con el estado mental del conductor, siendo mayor en caso de distracción [72]. Para la deducción del estado mental, se combinan varias métricas como el tiempo de reacción de frenado o el

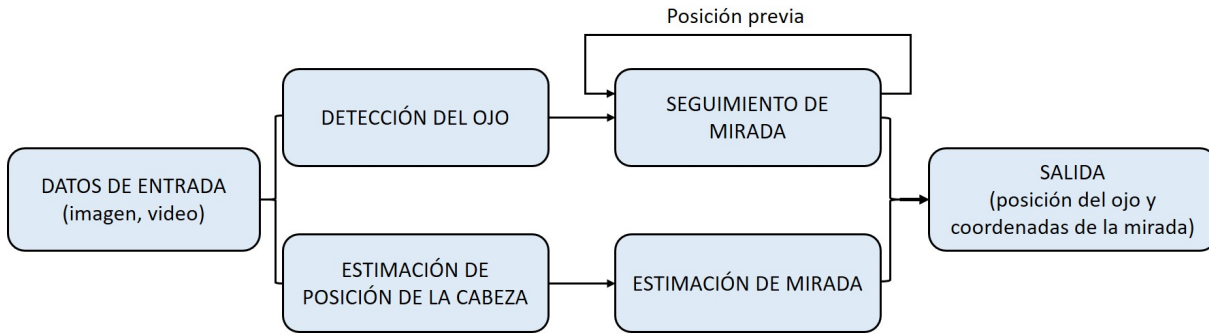


Figura 3.8: Representación genérica de las técnicas empleadas para el seguimiento del ojo y las coordenadas de la mirada [66].

tiempo de detección periférica. En caso de distracción, la *velocidad del vehículo* se ve alterada, aunque esta no presenta una correlación clara, pues puede aumentar o disminuir en función del escenario en particular. Por una parte, la reducción de velocidad es fruto de un mecanismo de compensación que se genera en el conductor, pues este trata de reducir la velocidad para minimizar así cualquier riesgo potencial. Sin embargo, otros estudios [73] muestran el aumento de velocidad como consecuencia del grado de aislamiento acústico que se consigue actualmente en el interior del vehículo, reduciendo la sensación de velocidad en el conductor. Por tanto, la variación de velocidad es un indicador de la distracción del conductor, pudiendo aumentar o disminuir en cada situación, en función de múltiples factores del entorno, vehículo y conductor.

Por ello, es de suma importancia conocer el estado de fatiga del conductor a través de la obtención de datos de su mirada para, así, prevenir situaciones que pueden comprometer la seguridad vial e intrínseca. En otras palabras, la implementación de técnicas centradas en el conductor, como el seguimiento de mirada, podría disminuir el número de accidentes de tráfico [74].

3.4.2. Técnicas de seguimiento de la mirada

En algunos ADAS actuales, la detección del estado del conductor se realiza mediante *eye tracking* (ver Figura 3.8). Para ello, la introducción de técnicas basadas en inteligencia artificial así como el desarrollo de la electrónica permiten el empleo de dispositivos no invasivos de seguimiento de la mirada (*eye trackers*) sobre la cabeza del conductor. Este tipo de dispositivos incluye las cámaras monoculares, estereoscópicas o infrarrojas. La detección del ojo y el seguimiento de la mirada en el entorno de conducción resulta complejo, pues hay numerosos factores que pueden dificultarlos, como la apertura del ojo, su tamaño, o la posición de la cabeza. Existen diversas técnicas para detectar la mirada [66], pudiendo estar basadas en su forma, características particulares, o apariencia.

Para cuantificar los cambios en la mirada de los conductores, se emplea frecuentemente el porcentaje del centro de carretera (*percent road center*, PRC). Se trata de una medida intuitiva que proporciona el porcentaje de fijaciones de la mirada que están ubicadas en el área central de la carretera durante un intervalo temporal establecido [75]. Esta métrica aumenta cuando la demanda cognitiva también lo hace [76, 77].

3. ADAS basados en visión

Medición	Tipo de distracción detectada			Ventajas	Inconvenientes
	Visual	Cognitiva	Ambas		
Comportamiento en la conducción	Sí	No	No	Indica la causa de distracción en la conducción.	Requiere informes subjetivos complementarios para la obtención de resultados con precisión elevada.
Parámetros físicos	Sí	Sí	No	Distingue los tipos de distracción.	Limitación para distinguir tipos combinados de distracción.
Parámetros biológicos	Sí	Sí	Sí	Mide distracción cognitiva y visual.	Invasivo.
Informe subjetivo	No	Sí	No	Distingue el mecanismo de distracción subyacente.	Requiere manejo por parte de un experto.
Medición híbrida	Sí	Sí	Sí	Presenta mayor precisión para identificar los tipos de distracción. Complementa las carencias que tienen otros métodos.	Necesita sincronizar los datos de múltiples fuentes con frecuencias de muestreo diferentes.

Tabla 3.1: Resumen de las técnicas de medición empleadas en la detección de la distracción del conductor en el vehículo [66].

Algunas características del movimiento de los ojos se emplean frecuentemente para la detección de somnolencia en el conductor [78]. Entre las más empleadas, destacan el porcentaje de cierre ocular, la amplitud de parpadeo, la frecuencia del parpadeo, o el retraso de la reapertura del párpado. El *porcentaje de cierre ocular* (PERCLOS) refleja la proporción del tiempo en el cual los ojos se encuentran cerrados entre un 70 % y un 80 %, durante el intervalo temporal de un minuto. La *amplitud del parpadeo* es el tiempo consumido desde el inicio hasta el final del movimiento del párpado, generalmente del orden de milisegundos. Se trata de una magnitud que varía ampliamente en función de la persona, requiriéndose así una calibración previa para poder aplicar correctamente la técnica de medición. Por otro lado, la *frecuencia de parpadeo* es el número de veces que se realiza este movimiento por minuto, y se asocia habitualmente al comienzo del sueño cuando alcanza frecuencias elevadas. Finalmente, el *retraso de la reapertura del párpado* mide el tiempo que transcurre desde el cierre completo de los párpados hasta el comienzo de su reapertura. Este tiempo varía en el orden de milisegundos y, cuando supera los 500 ms, pasa a considerarse *microsueño*, que puede ser causante de accidentes graves.

Tras el estudio de numerosas técnicas [79] para conocer y analizar el estado del conductor durante la conducción, se llega a la conclusión de que la fusión de varias técnicas permite generar una respuesta óptima en los sistemas de detección de fatiga del conductor [61]. Para ello, se combina la información obtenida de los parámetros del vehículo (velocidad, aceleración y dirección) y de los parámetros físicos y biológicos (frecuencia de parpadeo, amplitud, fijación del ojo, etc.) del conductor. Esto genera una nueva técnica híbrida (ver Tabla 3.1) con la que se obtiene un análisis del estado del conductor con mayor fiabilidad, y precisiones entre un 81 % [80] y un 93 % [81], reduciendo así el riesgo de accidentes.

3.4.3. Algoritmos de procesamiento de la mirada en ADAS

Existe un gran número de métricas que se obtienen a partir de los datos visuales del conductor y se aplican en ADAS, como la diferencia entre los valores mínimo y máximo de una distribución de datos con distintos valores, la desviación estándar de la misma, la duración de señal, la máxima diferencia entre dos valores consecutivos, la media, la mediana, o el valor máximo y el mínimo de los datos, entre otros. Para tratar estos volúmenes de datos, existen distintos algoritmos que

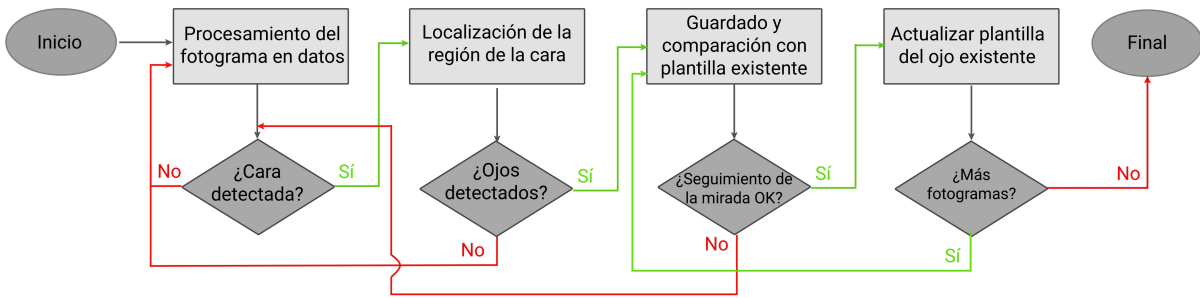


Figura 3.9: Diagrama de flujo de un algoritmo de *eye tracking* [66].

permiten modelar la información del ojo y la mirada con la finalidad de detectar el estado de alerta del conductor, así como su intención.

Estos algoritmos pueden emplear distintos métodos, entre los que destacan la lógica difusa [82], las técnicas de aprendizaje supervisado [83], no supervisado [84], o híbrido [85], las redes neuronales [86], o los modelos bayesianos [87], pudiendo ser incluso combinados entre sí para obtener así mejores resultados [88]. Cabe destacar que, a pesar de los múltiples algoritmos que existen, algunos de los empleados en *eye tracking* están basados en estrategias de aprendizaje profundo o *deep learning* (véase Capítulo 4). No obstante, la mayoría de estos algoritmos responden a una estructura común (ver Figura 3.9).

Por otra parte, la predicción de la intención del conductor [89] es una potencial funcionalidad de los ADAS que está siendo estudiada y desarrollada en la actualidad, pues permite predecir el comportamiento del conductor en el momento de realizar distintas maniobras, e incluso anticiparse a las mismas. Para ello, se emplean algoritmos basados en modelos estadísticos (como el modelo oculto de Markov [90]) que se aplican a los datos, obtenidos a partir de la secuencia de la mirada y de la posición de la cabeza del conductor.

En la Tabla 3.3 se presenta un resumen de distintos algoritmos que puede incluir un ADAS que utiliza datos de la mirada para la detección de fatiga o distracciones en el conductor, con la intención de que los elementos actuadores del sistema puedan alertarle.

3.5. Limitaciones

Una vez conocidas las técnicas de identificación del ojo y de seguimiento de mirada, los modelos y algoritmos que emplean, y sus ventajas e inconvenientes, conviene mencionar la importancia del desarrollo de los ADAS basados en estas técnicas, pues implican una mejora en el campo de la automoción, y más concretamente, en el contexto de la autonomía del vehículo. Sin embargo, la implementación de este tipo de sistemas se encuentra condicionada por factores como su coste, o su polivalencia y fiabilidad en distintos entornos de conducción. Por ello, Qasim Khan y Lee [66] realizaron un estudio de los requisitos fundamentales para una futura implementación a gran escala en el mercado automovilístico.

El factor económico es el principal factor que influye directamente sobre la aplicabilidad de estos sistemas, puesto que el precio de la tecnología de *eye tracking* es todavía elevado como

3. ADAS basados en visión

Algoritmo de detección de distracción/fatiga	Características monitorizadas	Algoritmo de detección de fatiga	Rendimiento
Teoría de la probabilidad (Red Bayesiana)	PERCLOS, cabeceo, orientación de cabeza, velocidad de parpadeo, dirección de la mirada, bostezos	Combinación de filtros Kalman promedados.	Muy elevado
Teoría de la probabilidad (Red Bayesiana)	PERCLOS, velocidad de parpadeo, dirección de la mirada, rotación de la cabeza.	Filtros adaptativos (Kalman).	Muy elevado
Sistema de conocimiento (sistema experto)	PERCLOS, velocidad de parpadeo, dirección de la mirada, cabeceo, orientación de la cabeza.	Filtros adaptativos (Kalman).	- - -
Sistema de conocimiento (autómata finito)	PERCLOS, velocidad de parpadeo, dirección de la mirada, bostezo, orientación de la cabeza.	Combinación de 4 métodos jerárquicos de seguimiento.	Estándar
Sistema de conocimiento (sistema experto)	PERCLOS, apertura de párpados, velocidad de parpadeo, orientación de la cabeza.	Ventana de búsqueda (basada en comparación con plantilla).	Elevado
Valor umbral	PERCLOS, velocidad de parpadeo, orientación de la cabeza.	Combinación de redes neuronales y algoritmo de condensación	Elevado
Valor umbral	PERCLOS, apertura de párpados.	Ventana de búsqueda (basada en comparación con plantilla).	Elevado
Valor umbral	Cierre continuo de los ojos.	Filtros adaptativos (Kalman)	Estándar
Valor umbral	Apertura de párpados.	Filtros adaptativos (Kalman)	Muy elevado

Tabla 3.3: Resumen de los distintos algoritmos de *eye tracking* más empleados [66].

para poder ser implementado en vehículos accesibles a un mayor segmento poblacional. Por ello, los ADAS implementados en gamas inferiores recurren al empleo de elementos más accesibles económicamente (como las cámaras monoculares), garantizando así precios más asequibles. Si bien, el software de detección debería estar adaptado para funcionar con imágenes de menor calidad.

Otro factor de gran relevancia es la flexibilidad, puesto que la mayoría de dispositivos de *eye tracking* requieren una calibración geométrica de las cámaras para cada conductor, y para cada uso, previa a su empleo, lo que supone un coste logístico añadido, por lo que un sistema agnóstico al conductor sería mucho más óptimo para ser instalado en un vehículo de uso frecuente.

La tolerancia de estos sistemas es, además, limitada, puesto que no funcionan correctamente en presencia de accesorios que interfieren en la detección del ojo o la mirada como son las gafas o lentillas. Por ello, el empleo de varias fuentes de luz para la detección de la mirada puede cubrir con relativa solvencia este tipo de limitaciones, evitándose así brillos, reflejos o cualquier otro tipo de interferencia [91]. También, una correcta modelización de los distintos tipos de lentes que puede portar el conductor bajo condiciones lumínicas variables puede ser de gran utilidad para el funcionamiento correcto del sistema en entornos con luminosidad cambiante en el exterior [92].

Adicionalmente, la interpretación de la relación entre el estado visual y cognitivo del conductor a través de la mirada puede ser un factor de gran relevancia, puesto que el análisis de los movimientos oculares se puede emplear para determinar el estado emocional del conductor, que a su vez presenta grandes influencias en su estado de atención, o en el modo de conducción. Sin embargo, se trata de un frente de investigación multidisciplinar con todavía un amplio margen de explotación y mejora.

Finalmente, el empleo de sistemas que se colocan sobre la cabeza del conductor resulta incómodo e intrusivo, pudiendo ser motivo de distracción. Por ello, se busca poder realizar el seguimiento de mirada a distancia, siempre que se garantice la misma precisión que se obtiene mediante un dispositivo no remoto [93].

Tras la observación de la importancia que tienen los ADAS basados en *eye tracking*, en este trabajo se propone una posible mejora para estos sistemas que, partiendo del conocimiento del comportamiento de la mirada del conductor, permita predecir su atención mediante el empleo de sensores más económicos y comunes en el mercado, como las cámaras monoculares entre otros.

4. Predicción de la atención en conducción

A lo largo de los capítulos anteriores, se ha podido observar que la mayoría de los ADAS desarrollan un mecanismo reactivo; es decir, su funcionamiento se rige por un cambio en los estímulos que reciben, que provoca una reacción al mismo. En este capítulo, se explora la propuesta de un sistema con funcionamiento predictivo, que sea capaz de predecir una situación de potencial riesgo, mediante la combinación de estímulos del interior y exterior del vehículo.

A lo largo del capítulo se emplea un gran volumen de terminología propia del campo de la Informática, y en concreto de la *Inteligencia Artificial* (IA). Por este motivo, se facilita el Anexo A en el que se descubren los principales conceptos de IA relevantes para este capítulo.

Particularmente, en este capítulo se exploran las bases del *aprendizaje automático*, así como conceptos relacionados con la atención del conductor, como la saliencia, o con la información visual del entorno, a través de la segmentación semántica, y se propone un modelo basado en los mismos que permite conocer las regiones a las que es más probable que el conductor dirija su atención. Además, se analizan las limitaciones del modelo, y se proponen mejoras con intención de generar resultados más precisos.

4.1. Motivación y contexto

En las últimas décadas, la *inteligencia artificial* (IA) ha cobrado gran importancia, permitiendo que los computadores actuales puedan llevar a cabo ciertas tareas que requieren un mínimo razonamiento sobre cierta información. Dentro de las técnicas de IA, destaca el denominado *machine learning*, donde el sistema no sólo razona, sino que es capaz de retroalimentarse y aprender conforme recibe información. Particularmente, son las llamadas *redes neuronales* las que, frente a un escenario concreto, aprenden a procesar información, análogamente a como lo haría una red neuronal biológica. Como modelo computacional, estas redes existen desde los años 50, pero empezaron a utilizarse ampliamente con el desarrollo y modernización de la tecnología disponible, así como con la mejora de la capacidad de cálculo de las computadoras. Así, se emplean en reconocimiento de voz o imagen, predicción de mercados, generación de texto, traducción de idiomas, prevención de fraudes, clasificación de alimentos a nivel industrial, conducción autónoma, análisis genéticos, pronósticos, etc.

El estudio de la atención humana también ha evolucionado gracias a la IA. Particularmente, múltiples trabajos han estudiado cómo modelar el comportamiento visual humano en diferentes entornos, incluyendo la conducción, donde se ha demostrado que el empleo de ADAS puede influir

4. Predicción de la atención en conducción

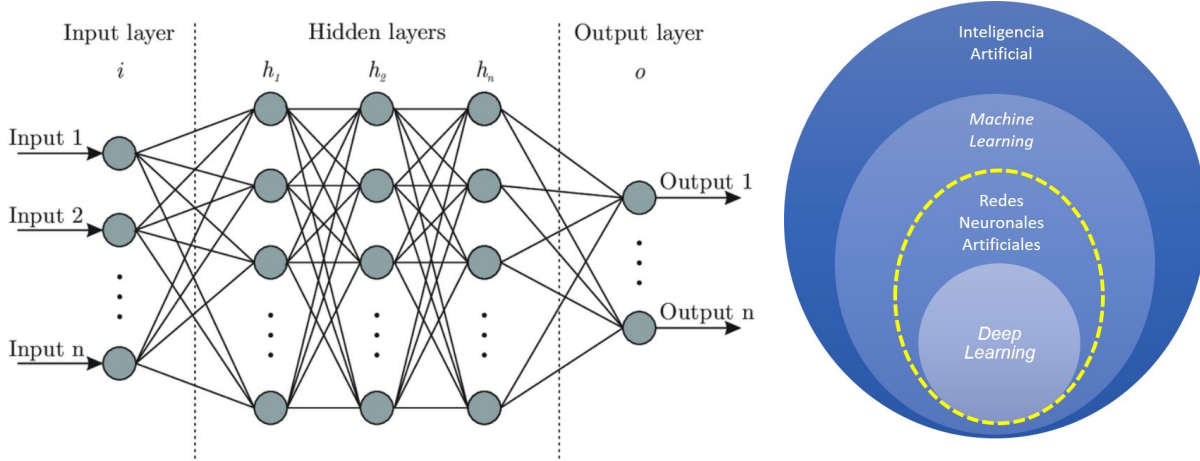


Figura 4.1: Izquierda: Estructura básica de una red neuronal [97]. Derecha: Adaptación del esquema general sobre la terminología que engloba la Inteligencia Artificial [94].

en su conducta [43]. Esto pone en evidencia la importancia del análisis del comportamiento del conductor a través de datos de la mirada.

4.1.1. Técnicas de *deep learning*

En los últimos años, la IA es empleada cada vez con más frecuencia en diversos entornos, pues ofrece un conjunto de técnicas que permiten crear sistemas inteligentes que trabajan y reaccionan como los seres humanos [94]. Entre las principales técnicas que engloba la IA, destacan los sistemas basados en conocimiento y el *machine learning*. Los *sistemas basados en conocimiento* (en inglés, *Knowledge-based Systems*, KBS), realizan un razonamiento simbólico similar al de los seres humanos para la resolución de problemas y el apoyo en la toma de decisiones, basándose en la evaluación de una serie de reglas. Un ejemplo de KBS son los denominados *sistemas expertos*, que imitan el razonamiento humano de un especialista a la hora de solucionar un problema específico [95].

Paralelamente, el amplio conjunto de técnicas de *aprendizaje automático* (en inglés, *machine learning*) se caracteriza por detectar patrones y extraer información inherente en los datos proporcionados, y aprender de forma automática, con la finalidad de cumplir un cierto objetivo [96].

En particular, las *redes neuronales artificiales* (en inglés, *Artificial Neural Networks*, NN) son técnicas de *machine learning* que emplean modelos computacionales estructurados por capas, que imitan la estructura y, tras su entrenamiento, el funcionamiento del cerebro humano. Las NN están compuestas por elementos, denominados *neuronas*, que se encuentran interconectados en una disposición o arquitectura determinada, y generan resultados que han sido procesados a partir de los datos aportados como entrada de la red (ver Figura 4.2a). Generalmente, el procesamiento de estos modelos se caracteriza por una estructura jerárquica, mediante representaciones de la información que varían de más simples a otras con mayor complejidad.

4. Predicción de la atención en conducción

Con los conceptos introducidos previamente, se define el *aprendizaje profundo* o *deep learning* como una sofisticación del *machine learning* que se caracteriza por incluir múltiples capas adicionales en las redes neuronales artificiales, obteniendo así mayores capacidades con respecto al modelo clásico [98].

Los modelos de *deep learning* están basados en modelos matemáticos complejos que *aprenden* de forma empírica con cada iteración de la información. Además, al obtener mejores resultados respecto a los obtenidos con modelos de NN tradicionales, los modelos basados en *deep learning* pueden realizar tareas comúnmente atribuidas al ser humano por ser de mayor complejidad, manteniendo un alto grado de precisión.

Uno de los principales campos que se encuentra en pujanza tecnológica durante los últimos años, y en los que se aplican estas técnicas es el ámbito de la *visión por computador*, que desarrolla métodos y procedimientos en forma de algoritmos para extraer información de las imágenes, pudiendo modificarlas, analizarlas, reconocer patrones, realizar mediciones, o incluso generar imágenes a partir de modelos. Para todo ello, se emplea un tipo de arquitecturas de NN conocido como *redes neuronales convolucionales* (en inglés, *Convolutional Neural Network*, CNN), que básicamente aplican filtros, extraen características y aprenden a predecir la respuesta esperada a partir de ellos (ver Anexo A).

En relación al contexto de este proyecto, cabe destacar la amplia aplicación que tienen las técnicas de *deep learning* en el campo de la automoción, pues permiten la implementación de CNN en ADAS con la finalidad de analizar y comprender el entorno del vehículo a través de la detección de señales de tráfico, la segmentación semántica, la predicción de profundidad de los elementos en el campo visual, o la predicción de atención del conductor, entre otros.

4.1.2. Saliencia

La información de la atención visual es de gran importancia en muchos campos, incluyendo la conducción [99]. Sin embargo, este tipo de información puede ser compleja de analizar, al involucrar múltiples fuentes de información, como los movimientos oculares o la orientación de la cabeza. Para facilitar el estudio de la atención, surgen diversas métricas. Este proyecto se centra en una de ellas, conocida como *saliencia visual*.

La *saliencia visual* es una métrica empleada para cuantificar la atención que presta un sujeto hacia un estímulo visual que destaca respecto a su entorno; o en otras palabras, cuánto atrae la atención del observador cada elemento de una escena. Se computa agregando información de la dirección de la mirada de múltiples usuarios, a lo largo del tiempo, y generalmente se representa mediante un mapa de calor probabilístico S , en el que cada elemento tiene asociado un valor normalizado $s \in [0, 1]$, donde las zonas que más atraen la atención (es decir, más salientes) poseen valores más cercanos a 1, y aquellas que generan menor interés visual, se encuentran próximas a 0.

En el estudio de la atención en la conducción realizado por Aletto et al. [100], se afirma que la *saliencia*, junto con otros subparámetros como la segmentación semántica o el reconocimiento del carril, puede tener un papel determinante en la conducción autónoma y asistida, aumentando así el interés en los últimos años de estudios sistemáticos del impacto de estos parámetros.

4. Predicción de la atención en conducción



Figura 4.2: Mapas de saliencia (superpuestos sobre su correspondiente imagen RGB) con información de la mirada del conductor obtenidos mediante la superposición de 25 fotogramas de secuencia de conducción [100]. Las regiones destacadas en colores cálidos en las imágenes indican las zonas que presentan mayor probabilidad de ser miradas por el conductor.

Uno de los principales factores que han impulsado este desarrollo es la capacidad computacional actual, que permite trabajar, analizar y modelar grandes volúmenes de datos, como podrían ser los de la atención en la conducción. Con este objetivo, surge el proyecto *Dr(eye)ve* [101], en el cual se basa este trabajo, y que mide y recopila información atencional de conductores bajo distintas condiciones (ver Anexo B).

Dada la importancia de los parámetros de la atención y la mirada del usuario, así como la actual carencia en la mayoría de ADAS en concepto de atención visual, en la Sección 4.2 se plantea el desarrollo de un sistema de predicción, en tiempo real, de la atención visual de los usuarios durante la conducción, que no requiere *eye tracking* constante del conductor. Por el contrario, este sistema aprende las características atencionales internas de los conductores, y su relación con la saliencia de la escena, de forma que es capaz de inferir dónde son más propensos a mirar los conductores en cada escenario. Las principales ventajas de esta propuesta incluyen la disminución de costes de inversión en sistemas que implican dispositivos tales como cámaras infrarrojas en el interior del vehículo, o gafas de *eye tracking*, simplificando en problema al uso de una cámara RGB orientada al exterior del vehículo. Además, la predicción en tiempo real podría ser empleada en un sistema de alerta al conductor, relativo a eventos que supongan riesgo potencial en la carretera y que, de otro modo, podrían pasar inadvertidos.

4.2. Desarrollo de un modelo genérico de predicción de saliencia

El análisis del estado del arte de ADAS en los capítulos anteriores ha permitido conocer las principales carencias que presentan estos sistemas. Por ello, se crea un modelo convolucional¹ que predice las zonas a las que es más probable que el conductor preste atención. Para afrontar este problema de forma óptima, se emplea una arquitectura *encoder-decoder*² (ver Figura 4.3), que facilita el análisis de imágenes mediante la extracción de características de las mismas a través

¹Toda la información relativa a Redes Convolucionales a la cual se hace referencia en este capítulo se encuentra en la Sección A.2 del Anexo A.

²Toda la información relativa al funcionamiento del modelo *encoder-decoder* a la que se hace referencia en esta sección se encuentra en Sección A.3 del Anexo A.

4. Predicción de la atención en conducción

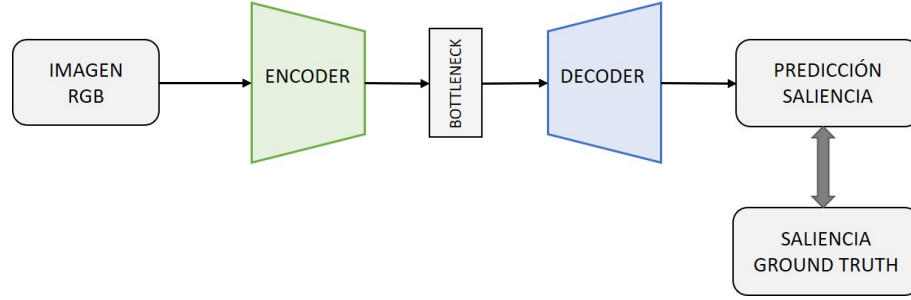


Figura 4.3: Esquema del modelo *encoder-decoder* empleado inicialmente, formado por capas convolucionales.

de *convoluciones*. La capacidad del modelo para resolver el problema depende notablemente de los *datos* empleados para su entrenamiento. Particularmente, es de gran importancia una correcta elección de la *función de pérdida*, pues es la que permite la correcta optimización del modelo. Tras la prueba del modelo optimizado, se calculan distintas *métricas* que proporcionan información cuantitativa sobre la precisión de los resultados obtenidos, y se generan múltiples resultados cualitativos para reforzarlos.

Inicialmente, se plantea una red con un *encoder* formado por cuatro capas, conteniendo cada una de ellas un bloque convolucional (convolución, normalización y función de activación). El *encoder* (ver Figura 4.4) recibe imágenes RGB, obtenidas mediante la extracción periódica de fotogramas a partir las secuencias de vídeo *ground truth* (empíricas) del *dataset* *Dr(eye)ve*³, y realiza un proceso de codificación o transformación de las mismas, tal y como se detalla en el Anexo A. Así, tras este proceso, la información codificada de la imagen entra en el cuello de botella, transformándose en el *vector latente*, que se emplea como entrada en el *decoder*.

El *decoder* (ver Figura 4.4) realiza el proceso contrario, transformando la información latente a través de cuatro capas, compuestas también por un bloque convolucional cada una de ellas, con la diferencia de que la salida obtenida es una imagen en blanco y negro. Esto implica que la convolución final está adaptada para que su salida sea una imagen con un único canal de valores en escala de grises.

Tras la preparación de los datos y la configuración de la arquitectura de la red, se realiza el entrenamiento y validación de la misma. En este proceso, la red es alimentada con todos los datos en cada iteración (*epoch*), sobre la cual se calcula el error de la misma en sus predicciones, a través de una *función de pérdida* (*loss function*). El objetivo final de la red es optimizarse para reducir su error en función de dichas pérdidas (para conocer más detalles, ver Subsección A.1.1 del Anexo A).

Particularmente, en este proyecto se calcula una función de pérdida de doble término. Por un lado, incluye la conocida como función L2, de mínimos cuadrados, o por sus siglas en inglés MSE (*mean squared error*), que mide el error píxel a píxel entre la predicción y el *ground truth*. Esta función es ampliamente empleada por su estabilidad, pues no presenta grandes desviaciones en el cálculo del error, y proporciona un único valor como solución [102]:

³El proceso se explica en el Anexo B.

4. Predicción de la atención en conducción

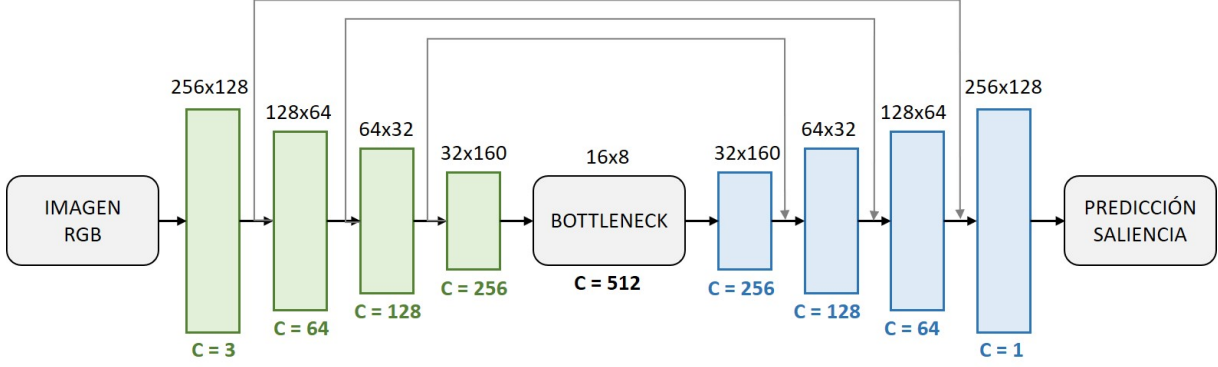


Figura 4.4: Esquema neuronal por capas del *encoder*, en verde, y *decoder*, en azul. Cada bloque representa un *bloque convolucional*, conformado por una operación de convolución, una normalización y una función de activación (ver Anexo A). Encima del bloque se detalla el tamaño de la entrada. Debajo del mismo, se indica el número de características (canales) que entran al bloque. Las líneas grises representan las *skip connections*, que conectan capas no consecutivas de la red, facilitando el proceso de decodificación de la información.

$$E_{L2} = \frac{1}{n} \sum_{i=1}^n (\hat{Y} - Y)^2 \quad (4.1)$$

donde \hat{Y} representa la predicción, e Y , el valor *ground truth*.

Por otro lado, y siguiendo la línea de otros trabajos relacionados del estado del arte (como el realizado por Aletto et al. [100]), se añade un segundo término basado en la divergencia de Kullback-Leibler (generalmente, KL ó KLDiv) [103] como función a optimizar, pues mide la diferencia entre dos distribuciones de probabilidad, P y Q (ver Ecuación 4.2).

$$D_{KL}(P \parallel Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)} \quad (4.2)$$

Una vez determinadas las funciones a optimizar, se lleva a cabo un proceso de entrenamiento, y se realiza un test con la versión mejor optimizada del modelo entrenado, obteniendo diversas métricas de relevancia en el contexto del problema: el coeficiente de correlación lineal, el valor de similitud, y la saliencia de la ruta de exploración normalizada.

El coeficiente de correlación lineal (en inglés, *linear correlation coefficient*, CC), o coeficiente lineal de Pearson, mide la relación estadística entre dos mapas de saliencia distintos, la predicción y el mapa *ground truth*, y proporciona un valor decimal en el rango $[-1, 1]$. Los extremos del intervalo denotan que los mapas guardan correlación, mientras que el valor 0 indica que los mapas no guardan ninguna correlación entre sí. Asimismo, el valor de similitud (*similarity score*, SIM) indica cuán parecidos son dos mapas de saliencia entre sí. No se trata de un valor normalizado, por lo que conforme más grande sea el valor, mayor será el parecido entre ambas imágenes. Por último, la saliencia de la ruta de exploración normalizada (*normalized scanpath saliency*, NSS) permite comparar predicciones de un modelo con las fijaciones reales de la mirada, proporcionando el valor medio del mapa de saliencia normalizado en aquellas zonas en las cuales hay fijaciones de la mirada. En este caso, cuanto mayor sea el parecido entre la saliencia

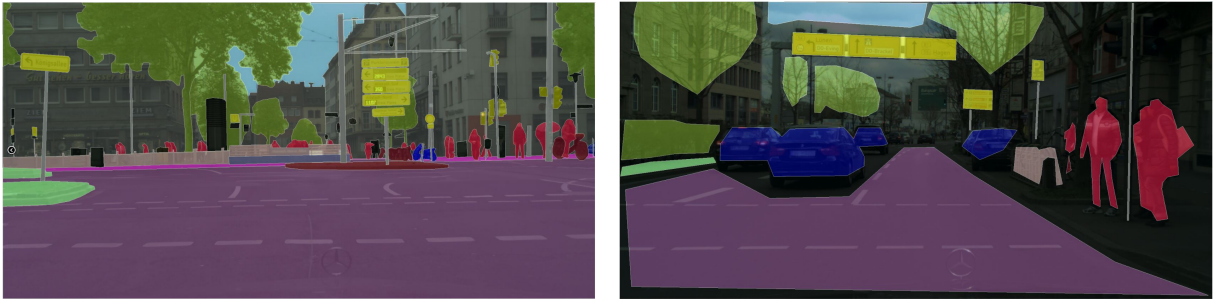


Figura 4.5: Clasificación de los distintos elementos de la imagen mediante segmentación semántica en las ciudades alemanas Düsseldorf (izquierda) y Dortmund (derecha) [105].

predicha y la *ground truth*, más elevada será la métrica.

Particularmente, se realiza un estudio de ablación en el que se analiza si la ponderación de la función KL puede tener cierto impacto en el rendimiento de la red. Tras analizar las métricas, se observa que la inclusión de KL las mejora, si bien su ponderación no resulta significativa.

Posteriormente, se decide realizar una segunda etapa de estudio de ablación para mejorar el rendimiento del modelo a través de la modificación de su arquitectura. Para ello, se tiene en cuenta la importancia de que un sistema aplicado en conducción ha de cumplir requisitos de seguridad exigentes. Así, se busca un modelo que permita detectar y reconocer situaciones potencialmente peligrosas en las que eventos puntuales e imprevistos, como un ciclista, un peatón, un animal, o cualquier objeto, puedan cruzarse en la trayectoria del vehículo de forma inesperada, pudiendo llegar a provocar un accidente. Por ello, se diseña un modelo con nociones de *segmentación semántica*.

4.2.1. Segmentación semántica

La segmentación semántica es una técnica que otorga una etiqueta o categoría a cada píxel de una imagen, con la finalidad de agruparlos, delimitando las regiones que representan una misma clase en la imagen. Se trata de una herramienta empleada en multitud de campos como la generación de imágenes médicas, la inspección industrial o la conducción autónoma, donde se emplea en la identificación de vehículos, peatones, señales de tráfico, aceras, y otros elementos de la calzada (ver Figura 4.5) [104].

La *segmentación semántica* permite la identificación de objetos en distintas áreas de la imagen, pudiendo incluso tener formas irregulares. Estas características suponen una gran ventaja frente a los sistemas de detección de objetos, en los cuales los objetos han de encajar en un cuadro delimitador (*bounding box*) [106] para ser reconocidos. Además, resultan idóneas para un problema en el que se pueden dar situaciones inesperadas en las que se produzca la aparición de objetos en el área del vehículo, pudiendo encontrarse en diversas posiciones y formas, e incluso pudiendo encontrarse deformadas por factores como la velocidad del vehículo o las condiciones lumínicas.

Para su inclusión en el modelo, se toma la arquitectura empleada, y se le añade una segunda rama formada por un modelo de segmentación semántica pre-entrenado, con su respectivo

4. Predicción de la atención en conducción

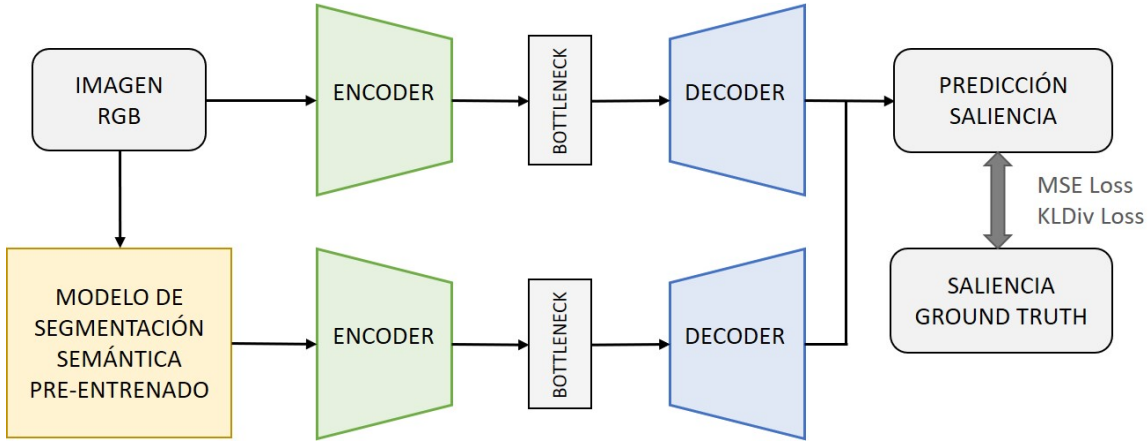


Figura 4.6: Esquema del modelo *encoder-decoder* de doble rama, en el cual se incluye un modelo de segmentación semántica pre-entrenado.

Métricas	Modelo inicial (sin segmentación)	Modelo con segmentación	Modelo con segmentación (KL ponderado)
CC ↑	0.34	0.35	0.40
NSS ↑	1.48	1.50	1.85
SIM ↑	0.28	0.29	0.28

Tabla 4.1: Estudio de ablación de la incorporación de la segmentación semántica, así como de la ponderación de la función de pérdida KL. Se observa que la adición de la segmentación semántica mejora las métricas obtenidas, así como la correcta ponderación de las funciones de pérdida.

encoder, cuello de botella y *decoder*.

Como se observa en la Figura 4.6, la entrada a la segunda rama es la misma que la de la rama original: las imágenes pretratadas, reescaladas y emparejadas con sus respectivas saliencias. Sin embargo, en esta nueva rama, las imágenes entran a un modelo de *segmentación semántica* pre-entrenado [107], *DeepLabV3*, que utiliza una red convolucional de clasificación, formada por cincuenta capas (llamada *ResNet-50*⁴), que segmenta la imagen en 21 categorías, generando imágenes de salida de *segmentación semántica* con 21 canales.

Los datos segmentados se introducen en un modelo *encoder-decoder* idéntico al de la rama principal, del cual se obtiene una segunda predicción de saliencia. Finalmente, se realiza una última convolución que combina las predicciones de ambas ramas (dos imágenes en blanco y negro), obteniendo una predicción conjunta de saliencia. En este caso, se calculan las funciones de pérdida L2 y KL a partir de la predicción conjunta, y se optimizan consecuentemente los pesos de las neuronas de ambas redes. De este modo, la red aprende, a través del entrenamiento, a valorar qué rama ha de tener más peso para realizar una predicción, o qué características de cada una son relevantes. Como se observa en la Tabla 4.1, se ha realizado el estudio de ablación análogo al previo, estudiando la importancia de la adición de la segmentación semántica al modelo, así como de una correcta ponderación de las funciones de pérdida, donde sí que se aprecia la mejora de las métricas en el caso de incluir tal modelo de *segmentación semántica* en la red combinado con una correcta ponderación.

⁴<https://pytorch.org/vision/stable/models.html#deeplabv3>

4. Predicción de la atención en conducción

Métricas	Modelo inicial	Modelo de doble rama (4 capas)	Modelo de doble rama (3 capas)
CC \uparrow	0.34	0.40	0.48
NSS \uparrow	1.48	1.85	2.22
SIM \uparrow	0.28	0.28	0.27

Tabla 4.2: Estudio de ablación que permite identificar el *overfitting* de la red. Cuando se reduce el número de capas, la red presenta una significativa mejora en las métricas.

No obstante, tras realizar el entrenamiento y el test, se observa que los resultados pueden no ajustarse correctamente al problema, siendo esto un indicio de que se está produciendo un sobreajuste (*overfitting*) (ver Anexo A) de la red. Las principales alternativas a este problema residen en la eliminación de capas de la red, el aumento del número de datos empleados, o el *dropout* o descarte aleatorio parcial de los datos en cada una de las capas. En este caso, se elimina una capa en el *encoder* y en el *decoder* respectivamente, lo cual permite trabajar con una red con menos parámetros, que en definitiva resulta en mejores métricas y predicciones más precisas (ver Tabla 4.2).

Sin embargo, al revisar los resultados cualitativos, se observa una gran variación de la precisión de predicción del modelo en función de las condiciones lumínicas y atmosféricas, pues en las situaciones de lluvia, la lente de la cámara se encuentra salpicada con gotas de agua, y de noche, no se aprecia gran parte del entorno del vehículo, o se producen destellos de faros y luces. Así, se llega a la deducción de que el modelo podría responder de forma distinta en función de estas condiciones, por lo que se procede a realizar un estudio de los distintos casos que se podrían dar, y se desarrollan modelos adaptados a cada una de dichas condiciones.

4.3. Desarrollo de modelos *específicos* de predicción de saliencia

Se realiza una división de los datos del *dataset* en función de las condiciones lumínicas y climáticas del momento, en tres escenarios: luz diurna, lluvia y luz nocturna. Las condiciones de luz diurna engloban días soleados así como nublados, sin presencia de precipitaciones. La selección de casos lluviosos engloba todos aquellos producidos en presencia de luz solar. Finalmente, se agrupan todos aquellos casos que se dan en un entorno nocturno. Con los datos divididos en las tres situaciones específicas, se entrena el modelo para cada caso, obteniendo así tres modelos entrenados y probados con datos de condiciones específicas. Con ello, se han calculado las métricas previamente comentadas (ver Tabla 4.3).

Al analizar las métricas obtenidas, se observa que para los casos lluvioso y nocturno, los valores han mejorado considerablemente, superando las cifras obtenidas en la versión previa del modelo (ver Tabla 4.3). Sin embargo, en el caso de las condiciones diurnas, se observa que los valores son inferiores. Esta situación puede deberse a la mayor variabilidad que experimentan los datos agrupados en condiciones diurnas, pues hay imágenes con gran luminosidad en las que se producen destellos, grandes contrastes, o acentuación de luces y sombras; o imágenes con el cielo nublado, generando mayor homogeneidad en la escena. Por tanto, estas variaciones de los parámetros de las imágenes podrían ser las causantes de una red poco ajustada al problema, al no haber podido aprender a identificar estas situaciones tan variables.

4. Predicción de la atención en conducción

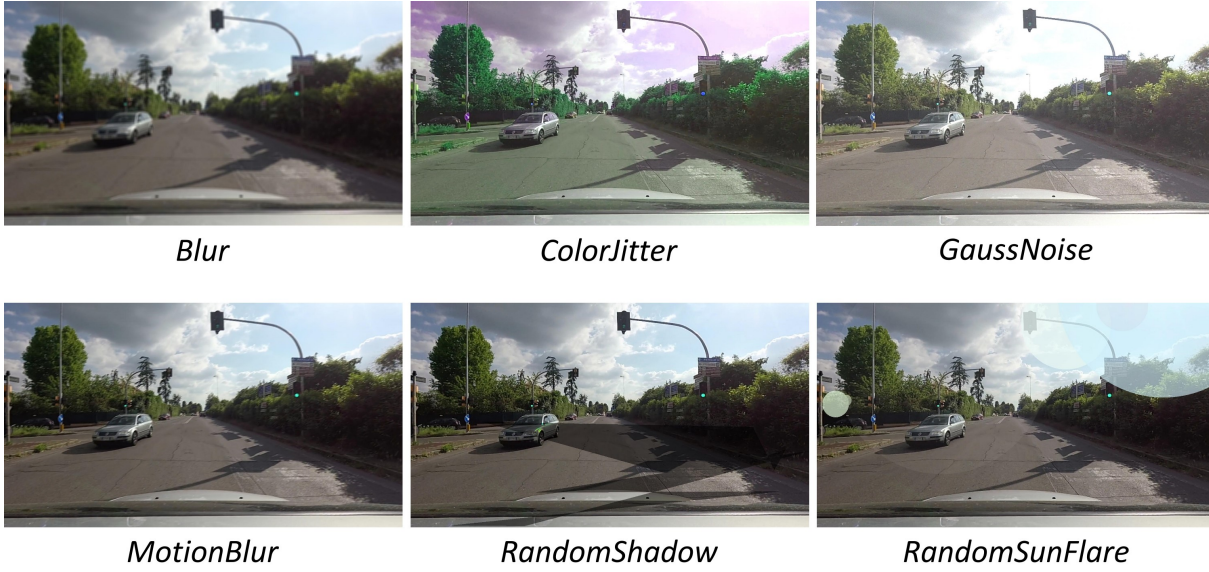


Figura 4.7: Ejemplo de las transformaciones realizadas sobre las imágenes del *dataset*, obtenidas de la librería *Albumentations*.

Métrica \ Modelo	Condiciones diurnas		Condiciones nocturnas		Condiciones de lluvia	
	Genérico	Específico	Genérico	Específico	Genérico	Específico
CC ↑	0.41	0.37	0.44	0.53	0.47	0.57
NSS ↑	2.11	1.95	2.21	2.34	2.47	2.57
SIM ↑	0.20	0.21	0.25	0.32	0.23	0.31

Tabla 4.3: Estudio de ablación en el que se comparan las métricas obtenidas con modelos entrenados en condiciones específicas, con los resultados de un modelo entrenado genéricamente. En dos de los tres casos, el modelo específico genera mejores cifras.

Por ello, se realiza un aumento de datos (comúnmente denominado mediante el término inglés *data augmentation*) mediante la librería *Albumentations*⁵, que incrementa la cantidad de imágenes del *dataset* mediante la aplicación de diferentes transformaciones a los datos. Particularmente, y dada la naturaleza de este trabajo, se han escogido transformaciones que modifican la luz, el contraste, o las sombras de la imagen aleatoriamente (ver Figura 4.7). Así, se contrarresta la variabilidad en los datos originales. El proceso completo y las transformaciones empleadas se explican en la Sección B.1.2 del el Anexo B.

Como se observa en la Tabla 4.4, las tres métricas obtenidas mejoran al realizar *data augmentation* sobre los datos en condiciones de luz diurna, superando incluso los valores del modelo genérico. Además, la observación cualitativa de los resultados obtenidos permite comprobar que se alcanza una saliencia más ajustada a los datos *ground truth* tras la realización de aumento de datos (ver Figura 4.8). Estos resultados confirman que la variabilidad en las imágenes empleadas requería un mayor volumen de datos con los que entrenar la red para su ajuste correcto. Así, se obtienen entonces los modelos optimizados en base a las tres condiciones planteadas y que, por tanto, generan mejores resultados.

⁵<https://albumentations.ai/docs/>

4. Predicción de la atención en conducción

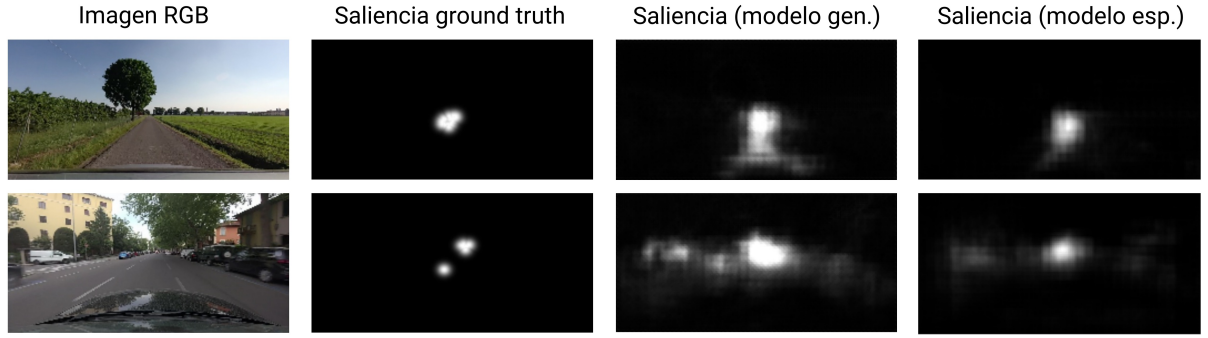


Figura 4.8: Para dos escenas distintas, se compara el resultado del modelo genérico frente al modelo específico para las condiciones de tales escenas.

Métrica \ Modelo	Condiciones diurnas			Condiciones nocturnas		Condiciones de lluvia	
	Genérico	Específico (no augm.)	Específico (augm.)	Genérico	Específico	Genérico	Específico
CC ↑	0.41	0.37	0.45	0.44	0.53	0.47	0.57
NSS ↑	2.11	1.95	2.20	2.21	2.34	2.47	2.57
SIM ↑	0.20	0.21	0.24	0.25	0.32	0.23	0.31

Tabla 4.4: Estudio de ablación en el que se analizan los resultados para el caso diurno. Se observa una mejora de las métricas al emplear aumento de datos, superando los valores obtenidos con anterioridad.

En la Figura 4.9 se pueden observar algunos de los resultados cualitativos obtenidos en los modelos mencionados. En el caso diurno con *data augmentation*, se observa que la predicción se aproxima correctamente a los datos *ground truth*, pues el modelo es capaz de inferir que la atención del conductor se dirige a aquellos vehículos que se cruzan en su trayectoria o se acercan a su vehículo. Además, el tercer ejemplo en condiciones de día presenta un patrón de la mirada con mayor amplitud en la dirección horizontal, pues la vista se compone de una carretera ancha con un vehículo en la lejanía, al cual el modelo ha sabido prestar atención, infiriendo así un comportamiento similar al de los datos empíricos.

Los resultados obtenidos en condiciones de lluvia presentan un ajuste muy acertado, pues el modelo es capaz de inferir las regiones de interés, demostrando que estas no siempre se encuentran centradas en la imagen, como ocurre en caso de curvas o en presencia de vehículos próximos. Con luz nocturna se produce un efecto de mirada al centro de la imagen, como se observa en el primer resultado, pues la luminosidad reducida de la escena no permite la visualización de todos los elementos que la componen. Sin embargo, se observa la correcta estimación de regiones salientes no centradas, provocadas por cambios en la dirección del vehículo o, nuevamente, la aproximación de otros vehículos.

4.4. Resumen

En este proyecto se ha creado en primer lugar un modelo *encoder-decoder* al que se le ha realizado un estudio de ablación, añadiendo la función de pérdida KL y ponderándola junto a la función L2, comprobándose que dicha combinación de funciones mejoraba las métricas obtenidas.

4. Predicción de la atención en conducción

A continuación, se ha desarrollado una red de doble rama, que mantiene la estructura *encoder-decoder* previa en la primera, e incorpora un modelo pre-entrenado de segmentación semántica seguido de un *encoder-decoder* en la segunda. Ambas estructuras son idénticas, exceptuando el modelo de segmentación, y producen dos predicciones de saliencia, que se combinan finalmente en una única predicción, que es más precisa que las anteriores.

Tras esta implementación, se decide estudiar el funcionamiento del modelo en tres situaciones lumínicas y ambientales distintas. Primero, se analizan tres modelos entrenados y probados específicamente para cada una de las tres situaciones particulares. También, se entrena un cuarto modelo con los datos de *train* de los tres casos, y se realiza un test específico para cada situación. Las métricas de los casos de noche y lluvia, mejoran. Sin embargo, con luz de día, son inferiores a las condiciones obtenidas en la versión previa. Por ello, se aplica *Data Augmentation* a los datos pertenecientes al caso de luz diurna, se entrena el modelo específico, y se prueba con los datos aumentados. Como se observa en la Tabla 4.4, se obtienen métricas superiores a las obtenidas anteriormente que, junto a la información que aportan los mapas de saliencia obtenidos (ver Figura 4.8), confirman que la variabilidad de las escenas bajo la luz diurna es mayor, y que el modelo necesita, por tanto, más datos para aprender a generar la respuesta esperada.

Por tanto, a lo largo de todo este proceso se aprende la importancia del diseño del modelo tratando de garantizar su adecuada adaptación al problema a resolver. En este proyecto, ha sido el caso de la incorporación de *segmentación semántica* al modelo previamente desarrollado. Además, la función objetivo posee un papel de gran relevancia en la optimización del modelo, lo que la hace fundamental en el proceso. Finalmente, es importante trabajar con los datos adecuados para la resolución del problema de forma eficaz, por lo que la obtención de un *dataset* acorde al problema resulta crucial

Actualmente, el sistema propuesto es capaz de inferir la región actual a la que es más probable que dirija su atención el conductor. Aunque su implementación en un ADAS no es trivial, da pie a múltiples aplicaciones que podrían mejorarlo. Por un lado, en presencia de un *eye tracker*, se podría alertar al usuario si su mirada no recae en las zonas salientes que deberían atraerla, reduciendo así el riesgo de distracción. Adicionalmente, si la escena no presenta grandes cambios en su saliencia, existe un amplio riesgo de distracción del conductor ante la falta de estímulos, por lo que se podrían poner en funcionamiento otros sensores que vigilaran su nivel de atención. Por otro lado, la saliencia puede reforzar procesos de reconocimiento de elementos que afectan directamente al proceso de conducción, pudiendo así ignorar aquellos que realmente no están dirigidos al conductor en dicha situación (por ejemplo, señales de vías cercanas por las que no se está circulando).

En definitiva, la propuesta de un sistema de predicción de saliencia permite una mayor comprensión de la escena, lo que puede ser de gran utilidad en aplicaciones futuras.

4. Predicción de la atención en conducción



Figura 4.9: Resultados obtenidos con los modelos optimizados en base a las tres condiciones lumínicas y meteorológicas planteadas.

5. Conclusiones

En este proyecto se ha llevado a cabo un estudio y análisis del estado del arte de los ADAS que se emplean actualmente en vehículos. Para ello, se ha analizado el campo de la conducción autónoma, conociendo los distintos niveles de automatización que pueden alcanzar los vehículos, así como los sistemas que los integran. Así, se ha observado que los vehículos actuales se encuentran entre los niveles SAE 2 y 3 de autonomía.

Seguidamente, se ha realizado un estudio y clasificación de los tipos de ADAS en función de los sensores que emplean, obteniendo así cinco clases distintas. Dentro de estas clases, se ha profundizado en los ADAS basados en *visión*, que emplean cámaras para la obtención de datos. Se ha estudiado el funcionamiento y características de estos sistemas, así como sus actuales limitaciones en la conducción.

Al realizar el estudio, se observa la importancia de la atención en la conducción para garantizar la seguridad vial, y evitar accidentes y situaciones de riesgo. Los sistemas que monitorizan al conductor poseen un importante papel en el análisis del estado y conducta de este, y su integración en el vehículo reduce el número de accidentes que se producen. Gran parte de estos sistemas funcionan con técnicas de *eye tracking*, también estudiadas en el proyecto. Sin embargo, se observa que estas presentan todavía algunas limitaciones, pues solo son accesibles a una minoría del segmento automovilístico relativa a vehículos de gama alta, requieren calibración previa, o pueden ser objeto de distracción, entre otras.

La mayoría de sistemas estudiados en estas revisiones son reactivos a estímulos, por lo que el desarrollo de un sistema predictivo podría influir positivamente en la conducción, pudiendo prever el comportamiento del conductor y alertar de posibles riesgos. Por ello, se propone un sistema que, a partir de la visualización del entorno exterior del vehículo (disponible en ADAS de gamas inferiores), permita predecir dónde mirará el conductor. Para ello, se crea una *red neuronal* que, a partir de un *dataset* compuesto por datos de la mirada (obtenidos con *eye tracking*) de distintos conductores en un entorno de conducción real, sepa indicar las regiones de mayor interés visual, a las que es más probable que el conductor preste atención.

Con intención de crear un modelo robusto, se añade al sistema información de segmentación semántica, que permite identificar elementos que puedan entrañar un potencial riesgo al cruzarse de forma inesperada con la trayectoria del vehículo. Para garantizar el buen funcionamiento del modelo, se hacen estudios cualitativos y cuantitativos que avalan las decisiones tomadas.

Sin embargo, debido a la amplitud de casos a los que se enfrenta el modelo en función de las condiciones ambientales y lumínicas que ocurren naturalmente en la conducción (situaciones nocturnas, cielos nublados, escenas con luz solar directa, etc), se estudia la respuesta del modelo

a través de su entrenamiento y test en situaciones distintas (luz diurna, luz nocturna, lluvia diurna), incluyendo también técnicas de aumento de datos adaptadas a escenarios de conducción.

El objetivo del proyecto era, en definitiva, analizar el estado del arte de la conducción autónoma, y potencialmente proponer mejoras para paliar limitaciones o carencias encontradas. Por ello, se ha realizado el estudio y análisis de los ADAS actuales, y se ha propuesto una técnica basada en seguimiento de la mirada capaz de predecir la atención del conductor y que, potencialmente, podría ser incluida en un ADAS.

5.1. Trabajo futuro

Con la elaboración de este trabajo de fin de grado, se han cumplido los objetivos propuestos inicialmente, pues se ha realizado un estudio de ADAS y una consecuente propuesta de aplicación de técnicas basadas en seguimiento de mirada. Sin embargo, a lo largo de todo el proyecto, han surgido nuevas ideas y posibles enfoques que podrían abordarse en el futuro:

- Prueba de nuevas parametrizaciones del modelo que desemboquen en una mejora del mismo, para la obtención de resultados más precisos, mediante el empleo de técnicas de *deep learning*.
- Estudio de la posibilidad de combinar sistemas basados en *visión* con sistemas que empleen otros sensores, como lidar o radar, creando así ADAS más robustos y con menos limitaciones, que puedan cubrir un mayor rango de situaciones.
- Además, al tratarse de sistemas que han de funcionar en tiempo real, su combinación con el análisis (también en tiempo real) del estado del conductor, ya fuese mediante *eye tracking* u otros parámetros fisiológicos, podría resultar de gran interés.
- Finalmente, el estudio de la potencial aplicación e integración del modelo propuesto como parte del ADAS de un vehículo.

5.2. Nota personal

Para mí, este proyecto ha supuesto una primera aproximación al campo de la investigación. Durante el curso 2019-2020, con la realización de una estancia en la *Technische Universität Wien*, pude descubrir que en los laboratorios y otras dependencias de una Universidad se llevan a cabo importantes tareas de investigación. Allí pude descubrir cómo se trabaja en un laboratorio de nanotecnología, se opera un Sistema Operativo Robótico, o se realizan simulaciones FEM, ensayos de seguridad, y análisis numéricos de termodinámica en fluidos. También fue en Viena donde me introduje en el mundo de la programación en *Python*, y donde escuché por primera vez el término *redes neuronales* en el contexto de la Industria 4.0.

Fueron estas experiencias las que me llevaron a plantearme la realización de un proyecto de carácter interdisciplinar, que permitiese mi acercamiento a otra materia como es la Informática, y su interconexión con el ámbito de la Ingeniería Mecánica en el cual he desarrollado mis estudios de

Grado. Así es como conocí al grupo de investigación *Graphics and Imaging Lab* de la Universidad de Zaragoza, donde Ingenieros de diversos ámbitos trabajan conjuntamente en un ambiente multidisciplinar, elaborando proyectos en el ámbito internacional. Particularmente, fueron Belén y Dani, mis directores, los que apostaron por el proyecto, y, posteriormente, lo apoyaron y orientaron con gran dedicación.

Este trabajo ha sido para mí una oportunidad de conocer y relacionar dos campos de conocimiento, desde una perspectiva investigadora, que me ha permitido conocer nuevas herramientas, métodos y formas de afrontar diversos problemas del ámbito de la Ingeniería.

Además, dado el interés que suscita esta línea de investigación, se planea continuar con este proyecto, con el objetivo de poder realizar una contribución en forma de publicación científica.

Bibliografía

- [1] How software is eating the car. <https://spectrum.ieee.org/cars-that-think/transportation/advanced-cars/software-eating-car>. [Último acceso 2021-06-10].
- [2] Tesla: Elon musk suggests autopilot not to blame for fatal crash. <https://www.bbc.com/news/technology-56799749>. [Último acceso 2021-06-09].
- [3] Tesla autopilot gets tricked into accelerating from 35 to 85 mph with modified speed limit sign. <https://electrek.co/2020/02/19/tesla-autopilot-tricked-accelerate-speed-limit-sign/>. [Último acceso 2021-06-10].
- [4] Vipin Kumar Kukkala, Jordan Tunnell, Sudeep Pasricha, and Thomas Bradley. Advanced driver-assistance systems: A path toward autonomous vehicles. *IEEE Consumer Electronics Magazine*, 7(5):18–25, 2018.
- [5] Cs230 deep learning. <https://cs230.stanford.edu/syllabus/>. [Último acceso 2021-01-29].
- [6] Daniel Martin, Ana Serrano, and Belen Masia. Panoramic convolutions for 360° single-image saliency prediction. In *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*, 2020.
- [7] Raimondas Zemblys, Diederick C Niehorster, and Kenneth Holmqvist. gazeNet: End-to-end eye-movement event detection with deep neural networks. *Behavior research methods*, 51(2):840–864, 2019.
- [8] Understanding pytorch with an example. <https://towardsdatascience.com/understanding-pytorch-with-an-example-a-step-by-step-tutorial-81fc5f8c4e8e>. [Último acceso 2021-01-12].
- [9] Handwritten digit recognition using pytorch - intro to neural networks. <https://towardsdatascience.com/handwritten-digit-mnist-pytorch-977b5338e627>. [Último acceso 2021-01-21].
- [10] Pytorch documentation. <https://pytorch.org/docs/stable/index.html>. [Último acceso 2021-06-07].
- [11] Albumentations documentation. <https://albumentations.ai/docs/>. [Último acceso 2021-06-05].

- [12] Jorge Godoy, Antonio Artuñedo, Rodolfo Haber, and Carlos González. Conducción autónoma y cooperativa—el programa autopía en España. In *XV Congreso Español sobre Sistemas Inteligentes de Transporte. Madrid: ITS Spain*, 2015.
- [13] SAE On-Road Automated Vehicle Standards Committee et al. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE Standard J*, 3016:1–16, 2014.
- [14] SAE Taxonomy. Definitions for terms related to driving automation systems for on-road motor vehicles (j3016). Technical report, Technical report, Society for Automotive Engineering, 2016.
- [15] SAE International. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *SAE International*, 4970(724):1–5, 2018.
- [16] Christina Rödel, Susanne Stadler, Alexander Meschtscherjakov, and Manfred Tscheligi. Towards autonomous cars: the effect of autonomy levels on acceptance and user experience. In *Proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications*, pages 1–8, 2014.
- [17] De 0 a 5: cuáles son los diferentes niveles de conducción autónoma, a fondo. <https://www.xataka.com/automovil/de-0-a-5-cuales-son-los-diferentes-niveles-de-conduccion-autonoma>. [Último acceso 2021-05-21].
- [18] Conducción autónoma — Los cinco niveles. <https://www.km77.com/reportajes/varios/conduccion-autonoma-niveles>. [Último acceso 2021-05-21].
- [19] Los niveles de la conducción autónoma. <https://www.cea-online.es/blog/213-los-niveles-de-la-conduccion-autonoma>. [Último acceso 2021-05-21].
- [20] Coche autónomo: los seis niveles de conducción (cuáles son, cómo funcionan, cuándo llegarán...) – Autobild.es. <https://www.autobild.es/practicos/coche-autonomo-seis-niveles-conduccion-cuales-son-como-funcionan-cuando-llegaran-490005>. [Último acceso 2021-05-21].
- [21] Francesco Walker, Anika Boelhouwer, Tom Alkim, Willem B Verwey, and Marieke H Martens. Changes in trust after driving level 2 automated cars. *Journal of advanced transportation*, 2018, 2018.
- [22] Juan Emilio Ballesteros. El espejo del éxito: Motor. *Cambio 16*, (2232):98–98, 2016.
- [23] Shantanu Ingle and Madhuri Phute. Tesla autopilot: semi autonomous driving, an up-tick for future autonomy. *International Research Journal of Engineering and Technology*, 3(9):369–372, 2016.
- [24] Sharon L Poczter and Luka M Jankovic. The google car: driving toward a better future? *Journal of Business Case Studies (JBBS)*, 10(1):7–14, 2014.
- [25] Douglas Kirsch. Autopilot and algorithms: accidents, errors, and the current need for human oversight. *Journal of clinical sleep medicine*, 16(10):1651–1652, 2020.

- [26] Probamos el Autopilot en Madrid ¿Podrá un Tesla Model 3 completar la Gran vía, Velázquez y Serrano de forma autónoma? (Vídeo) — forococheselectricos. <https://forococheselectricos.com/2021/04/probamos-el-autopilot-en-madrid-podra-un-tesla-model-3-completar-la-gran-via-velazquez-y-serrano-de-forma-autonoma-video.html>. [Último acceso 2021-05-21].
- [27] Super Cruise - Hands Free Driving — Cadillac Ownership. <https://www.cadillac.com/world-of-cadillac/innovation/super-cruise>. [Último acceso 2021-05-21].
- [28] Adam Ziebinski, Rafal Cupek, Damian Grzechca, and Lukas Chruszczyk. Review of advanced driver assistance systems (adas). In *AIP Conference Proceedings*, volume 1906, page 120002. AIP Publishing LLC, 2017.
- [29] W Cunningham. Us requiring back-up cameras in cars by 2018. road show by cnet (2014).
- [30] Sebastian Thrun, Mike Montemerlo, Hendrik Dohlkamp, David Stavens, Andrei Aron, James Diebel, Philip Fong, John Gale, Morgan Halpenny, Gabriel Hoffmann, et al. Stanley: The robot that won the darpa grand challenge. *Journal of field Robotics*, 23(9):661–692, 2006.
- [31] Erico Guizzo. How google’s self-driving car works.” iee spectrum online. *October*, 2011.
- [32] Nalin Jayaweera, Nandana Rajatheva, and Matti Latva-aho. Autonomous driving without a burden: View from outside with elevated lidar. In *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, pages 1–7. IEEE, 2019.
- [33] RADAR, Camera, LiDAR and V2X for Autonomous Cars — NXP Semiconductors. <https://www.nxp.com/company/blog/radar-camera-lidar-and-v2x-for-autonomous-cars:BL-RADAR-LIDAR-V2X-AUTONOMOUS-CARS>. [Último acceso 2021-05-20].
- [34] Charles Reinholtz, Thomas Alberi, David Anderson, Andrew Bacha, Cheryl Bauman, Stephen Cacciola, Patrick Currier, Aaron Dalton, Jesse Farmer, Ruel Faruque, et al. Darpa urban challenge technical paper. 2007.
- [35] Saúl Cova-Rocamora. Navegación y conducción autónoma de vehículos con geometría ackermann. 2019.
- [36] Javier Izquierdo-Reyes, Luis A Curiel-Ramirez, Ricardo A Ramirez-Mendoza, and M Rogelio Bustamante-Bello. A logic architecture for 360 adas-alerts for hazards detection based in driver actions. In *MATEC Web of Conferences*, volume 124, page 04011. EDP Sciences, 2017.
- [37] Robert Leibinger. Software architectures for advanced driver assistance systems (adas). *Agenda: Short overview of Elektrobite automotive*, 2015.
- [38] Javier Izquierdo-Reyes, Ricardo A Ramirez-Mendoza, and Martin R Bustamante-Bello. A study of the effects of advanced driver assistance systems alerts on driver performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 12(1):263–272, 2018.

- [39] Yoshihiko Takahashi, Tetsuya Komoguchi, Masato Seki, Nimesh Patel, David Auner, and Bruce Campbell. Alert method for rear cross traffic alert system in north america. Technical report, SAE Technical Paper, 2013.
- [40] Adam Ziebinski, Rafal Cupek, Damian Grzechca, and Lukas Chruszczyk. Review of advanced driver assistance systems (adas). In *AIP Conference Proceedings*, volume 1906, page 120002. AIP Publishing LLC, 2017.
- [41] Javier Izquierdo-Reyes, Luis A Curiel-Ramirez, Ricardo A Ramirez-Mendoza, and M Rogelio Bustamante-Bello. A logic architecture for 360 adas-alerts for hazards detection based in driver actions. In *MATEC Web of Conferences*, volume 124, page 04011. EDP Sciences, 2017.
- [42] SM Sarala, DH Sharath Yadav, and Asadullah Ansari. Emotionally adaptive driver voice alert system for advanced driver assistance system (adas) applications. In *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 509–512. IEEE, 2018.
- [43] John M Sullivan, Michael J Flannagan, Anuj K Pradhan, and Shan Bao. Literature review of behavioral adaptations to advanced driver assistance systems. 2016.
- [44] Javier Izquierdo-Reyes, Ricardo A Ramirez-Mendoza, and Martin R Bustamante-Bello. A study of the effects of advanced driver assistance systems alerts on driver performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 12(1):263–272, 2018.
- [45] Itay Gat, Meny Benady, and Amnon Shashua. A monocular vision advance warning system for the automotive aftermarket. *SAE transactions*, pages 403–410, 2005.
- [46] Philippe Morignot, Joshué Pérez Rastelli, and Fawzi Nashashibi. Arbitration for balancing control between the driver and adas systems in an automated vehicle: Survey and approach. In *2014 IEEE Intelligent Vehicles Symposium Proceedings*, pages 575–580. IEEE, 2014.
- [47] Bjorn Barrois, Stela Hristova, Christian Wohler, Franz Kummert, and Christoph Hermes. 3d pose estimation of vehicles using a stereo camera. In *2009 IEEE Intelligent Vehicles Symposium*, pages 267–272, 2009.
- [48] ¿Qué se esconde hoy en día detrás del retrovisor central? <https://www.motorpasion.com/espaciotoyota/que-se-esconde-hoy-en-dia-detras-del-retrovisor-central>. [Último acceso 2021-05-27].
- [49] Stereo camera setup used by vehicle manufacture Subaru. https://www.researchgate.net/figure/Stereo-camera-setup-used-by-vehicle-manufacture-Subaru-3_fig3_333419301. [Último acceso 2021-05-27].
- [50] East Autos: Cadillac represents Super Cruise control. <https://www.eastautos.co.uk/cadillac-super-cruise/>. [Último acceso 2021-05-29].
- [51] Yoichiro Iwasaki, Masato Misumi, and Toshiyuki Nakamiya. Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring. *Sensors*, 13(6):7756–7773, 2013.

- [52] Aneesh Paul, Rohan Chauhan, Rituraj Srivastava, and Mriganka Baruah. Advanced driver assistance systems. Technical report, SAE Technical Paper, 2016.
- [53] Jia Li and Zengfu Wang. Real-time traffic sign recognition based on efficient cnns in the wild. *IEEE Transactions on Intelligent Transportation Systems*, 20(3):975–984, 2018.
- [54] David Geronimo, Antonio M Lopez, Angel D Sappa, and Thorsten Graf. Survey of pedestrian detection for advanced driver assistance systems. *IEEE transactions on pattern analysis and machine intelligence*, 32(7):1239–1258, 2009.
- [55] Jianqing Zhu, Shengcai Liao, Dong Yi, Zhen Lei, and Stan Z Li. Multi-label cnn based pedestrian attribute learning for soft biometrics. In *2015 International Conference on Biometrics (ICB)*, pages 535–540. IEEE, 2015.
- [56] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.
- [57] Xuqin Yan and Yanqiang Li. A method of lane edge detection based on canny algorithm. In *2017 Chinese Automation Congress (CAC)*, pages 2120–2124. IEEE, 2017.
- [58] Chan Yee Low, Hairi Zamzuri, and Saiful Amri Mazlan. Simple robust road lane detection algorithm. In *2014 5th International Conference on Intelligent and Advanced Systems (ICIAS)*, pages 1–4. IEEE, 2014.
- [59] Simple lane detection with OpenCV. <https://medium.com/@mrhwick/simple-lane-detection-with-opencv-bfeb6ae54ec0>. [Último acceso 2021-06-04].
- [60] Sheila G Klauer, Thomas A Dingus, Vicki L Neale, Jeremy D Sudweeks, David J Ramsey, et al. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. 2006.
- [61] Mario I Chacon-Murguia and Claudia Prieto-Resendiz. Detecting driver drowsiness: A survey of system designs and technology. *IEEE Consumer Electronics Magazine*, 4(4):107–119, 2015.
- [62] Sotirios Diamantas, Stefanos Astaras, and Aristodemos Pnevmatikakis. Depth estimation in still images and videos using a motionless monocular camera. In *2016 IEEE international conference on imaging systems and techniques (IST)*, pages 129–134. IEEE, 2016.
- [63] Jamal Saboune, Mehdi Arezoomand, Luc Martel, and Robert Laganier. A visual blindspot monitoring system for safe lane changes. In *International Conference on Image Analysis and Processing*, pages 1–10. Springer, 2011.
- [64] Arun Tigadi, Rudrappa Gujanatti, Anil Gonchi, and Belagavi Klemsscet. Advanced driver assistance systems. *Int. J. Eng. Res. Gen. Sci*, 4(3):151–158, 2016.
- [65] Geoffrey Underwood, Katherine Humphrey, and Editha Van Loon. Decisions about objects in real-world scenes are influenced by visual saliency before and during their inspection. *Vision research*, 51(18):2031–2038, 2011.
- [66] Muhammad Qasim Khan and Sukhan Lee. Gaze and eye tracking: Techniques and applications in adas. *Sensors*, 19(24):5540, 2019.

- [67] Clara Marina Martinez, Mira Heucke, Fei-Yue Wang, Bo Gao, and Dongpu Cao. Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 19(3):666–676, 2017.
- [68] National Highway Traffic Safety Administration. Overview of motor vehicle crashes in 2019. *Traffic Safety Facts Research Note*, Report No. DOT HS 813 060, 2020, December.
- [69] National Highway Traffic Safety Administration. Preview of motor vehicle traffic fatalities in 2019. *Research Note*, Report No. DOT HS 813 021, 2020, October.
- [70] Thomas A Ranney, W Riley Garrott, and Michael J Goodman. Nhtsa driver distraction research: Past, present, and future. Technical report, SAE Technical Paper, 2001.
- [71] Yulan Liang and John D Lee. Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention*, 42(3):881–890, 2010.
- [72] David L Strayer, Joel M Cooper, Jonna Turrill, James Coleman, Nate Medeiros-Ward, and Francesco Biondi. Measuring cognitive distraction in the automobile. 2013.
- [73] Kristie L Young, Paul M Salmon, and Miranda Cornelissen. Distraction-induced driving error: An on-road examination of the errors made by distracted and undistracted drivers. *Accident Analysis & Prevention*, 58:218–225, 2013.
- [74] Hardeep Singh, Jagjit Singh Bhatia, and Jasbir Kaur. Eye tracking based driver fatigue monitoring and warning system. In *India International Conference on Power Electronics 2010 (IICPE2010)*, pages 1–6. IEEE, 2011.
- [75] Shiyang Yang, Jonny Kuo, and Michael G Lenné. Analysis of gaze behavior to measure cognitive distraction in real-world driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 1944–1948. SAGE Publications Sage CA: Los Angeles, CA, 2018.
- [76] Bryan Reimer, Bruce Mehler, Ying Wang, and Joseph F Coughlin. A field study on the impact of variations in short-term memory demands on drivers’ visual attention and driving performance across three age groups. *Human factors*, 54(3):454–468, 2012.
- [77] Joanne L Harbluk, Y Ian Noy, Patricia L Trbovich, and Moshe Eizenman. An on-road assessment of cognitive distraction: Impacts on drivers’ visual behavior and braking performance. *Accident Analysis & Prevention*, 39(2):372–379, 2007.
- [78] Erik J Sirevaag and John A Stern. Ocular measures of fatigue and cognitive factors. *Engineering psychophysiology: Issues and applications*, pages 269–287, 2000.
- [79] Ying Wang, Bryan Reimer, Jonathan Dobres, and Bruce Mehler. The sensitivity of different methodologies for characterizing drivers’ gaze concentration under increased cognitive demand. *Transportation research part F: traffic psychology and behaviour*, 26:227–237, 2014.
- [80] Yulan Liang, Michelle L Reyes, and John D Lee. Real-time detection of driver cognitive distraction using support vector machines. *IEEE transactions on intelligent transportation systems*, 8(2):340–350, 2007.

- [81] Masahiro Miyaji, Haruki Kawanaka, and Koji Oguri. Driver’s cognitive distraction detection using physiological features by the adaboost. In *2009 12th International IEEE conference on intelligent transportation systems*, pages 1–6. IEEE, 2009.
- [82] Matti Kutila, Pasi Pyykönen, Paul van Koningsbruggen, Nereo Pallaro, and Joshué Pérez-Rastelli. The deserve project: Towards future adas functions. In *2014 International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS XIV)*, pages 308–313. IEEE, 2014.
- [83] Fabio Tango and Marco Botta. Real-time detection system of driver distraction using machine learning. *IEEE Transactions on Intelligent Transportation Systems*, 14(2):894–905, 2013.
- [84] Ashesh Jain, Hema S Koppula, Shane Soh, Bharad Raghavan, Avi Singh, and Ashutosh Saxena. Brain4cars: Car that knows before you do via sensory-fusion deep learning architecture. *arXiv preprint arXiv:1601.00740*, 2016.
- [85] Tianchi Liu, Yan Yang, Guang-Bin Huang, Yong Kiang Yeo, and Zhiping Lin. Driver distraction detection using semi-supervised machine learning. *IEEE transactions on intelligent transportation systems*, 17(4):1108–1120, 2015.
- [86] Zhitao Xiao, Zhiqiang Hu, Lei Geng, Fang Zhang, Jun Wu, and Yuelong Li. Fatigue driving recognition network: Fatigue driving recognition via convolutional neural network and long short-term memory units. *IET Intelligent Transport Systems*, 13(9):1410–1416, 2019.
- [87] Qiang Ji, Zhiwei Zhu, and Peilin Lan. Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE transactions on vehicular technology*, 53(4):1052–1068, 2004.
- [88] Philippe Morignot, Joshué Pérez Rastelli, and Fawzi Nashashibi. Arbitration for balancing control between the driver and adas systems in an automated vehicle: Survey and approach. In *2014 IEEE Intelligent Vehicles Symposium Proceedings*, pages 575–580. IEEE, 2014.
- [89] Brendan Morris, Anup Doshi, and Mohan Trivedi. Lane change intent prediction for driver assistance: On-road design and evaluation. In *2011 IEEE Intelligent Vehicles Symposium (IV)*, pages 895–901. IEEE, 2011.
- [90] Seifemichael B Amsalu, Abdollah Homaifar, Ali Karimoddini, and Arda Kurt. Driver intention estimation via discrete hidden markov model. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2712–2717. IEEE, 2017.
- [91] Xia Liu, Fengliang Xu, and Kikuo Fujimura. Real-time eye detection and tracking for driver observation under various light conditions. In *Intelligent Vehicle Symposium, 2002. IEEE*, volume 2, pages 344–351. IEEE, 2002.
- [92] Walter J Talamonti Jr, Wenyan Huang, Louis Tijerina, and Dev Kochhar. Eye glance and head turn correspondence during secondary task performance in simulator driving. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 57, pages 1968–1972. SAGE Publications Sage CA: Los Angeles, CA, 2013.

- [93] Dan Su, You-Fu Li, and Hao Chen. Toward precise gaze estimation for mobile head-mounted gaze tracking systems. *IEEE Transactions on Industrial Informatics*, 15(5):2660–2672, 2018.
- [94] Bonifacio Martín Del Brío and Alfredo Sanz Molina. *Redes neuronales y sistemas borrosos*. Ra-ma Madrid, 2001.
- [95] Sebastián Badaró, Leonardo Javier Ibañez, and Martín Jorge Agüero. Sistemas expertos: fundamentos, metodologías y aplicaciones. *Ciencia y tecnología*, (13):349–364, 2013.
- [96] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.
- [97] Artificial Neural Network Architecture. <https://www.researchgate.net/profile/Facundo-Bre/publication/321259051/figure/fig1/AS:61432925049652901523478915726/Artificial-neural-network-architecture-ANN-i-h-1-h-2-h-n-o.png>. [Último acceso 2021-06-08].
- [98] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [99] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. Saliency in vr: How do people explore virtual environments? *IEEE transactions on visualization and computer graphics*, 24(4):1633–1642, 2018.
- [100] Stefano Alletto, Andrea Palazzi, Francesco Solera, Simone Calderara, and Rita Cucchiara. Dr (eye) ve: a dataset for attention-based tasks with applications to autonomous and assisted driving. In *Proceedings of the ieee conference on computer vision and pattern recognition workshops*, pages 54–60, 2016.
- [101] Andrea Palazzi, Davide Abati, Simone Calderara, Francesco Solera, and Rita Cucchiara. Predicting the driver’s focus of attention: the dr(eye)ve project. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
- [102] Differences between L1 and L2 as Loss Function and Regulariation. <http://www.chioka.in/differences-between-l1-and-l2-as-loss-function-and-regularization/>. [Último acceso 2021-06-13].
- [103] Fernando Pérez-Cruz. Kullback-leibler divergence estimation of continuous distributions. In *2008 IEEE international symposium on information theory*, pages 1666–1670. IEEE, 2008.
- [104] Image Processing and Computer Vision - ¿Qué es la segmentación semántica? <https://es.mathworks.com/solutions/image-video-processing/semantic-segmentation.html>. [Último acceso 2021-06-13].
- [105] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

- [106] A gentle introduction to Object Recognition with Deep Learning. <https://machinelearningmastery.com/object-recognition-with-deep-learning/>. [Último acceso 2021-06-13].
- [107] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [108] João Carlos Figueira Pujol and Joao Mario Andrade Pinto. A neural network approach to fatigue life prediction. *International Journal of Fatigue*, 33(3):313–322, 2011.
- [109] El Legado Cajal. Instituto Cajal. CSIC. <http://www.cajal.csic.es/legado.html>. [Último acceso 2021-06-10].
- [110] Santiago Ramón y Cajal: *Recuerdos de mi vida*. https://cvc.cervantes.es/ciencia/cajal/cajal_recuerdos/recuerdos/labor_05.htm. [Último acceso 2021-06-10].
- [111] Nelson Marcelo Romero Aquino and Eustaquio Alcides Martínez Jara. Aplicación de redes neuronales artificiales en la orientación vocacional. 2014.
- [112] Qué es overfitting y underfitting y cómo solucionarlo. <https://www.aprendemachinelearning.com/que-es-overfitting-y-underfitting-y-como-solucionarlo/>. [Último acceso 2021-06-10].
- [113] Boris Hanin and David Rolnick. How to start training: The effect of initialization and architecture. *arXiv preprint arXiv:1803.01719*, 2018.
- [114] Convolutional Neural Networks for Visual Recognition. <https://cs231n.github.io/convolutional-networks/>. [Último acceso 2021-06-11].
- [115] Manuel Lagunas Arto and Elena Garcés García. A Computational Model of Icons Appearance. 2017.
- [116] A comprehensive guide to Convolutional Neural Networks. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>. [Último acceso 2021-06-11].
- [117] Understanding Encoder-Decoder Sequence to Sequence Model. <https://towardsdatascience.com/understanding-encoder-decoder-sequence-to-sequence-model-679e04af4346>. [Último acceso 2021-06-11].
- [118] What is an encoder decoder model? <https://towardsdatascience.com/what-is-an-encoder-decoder-model-86b3d57c5e1a>. [Último acceso 2021-06-11].
- [119] AImageLab Laboratory. <https://aimagelab.ing.unimore.it/imagelab/page.asp?IdPage=8>. [Último acceso 2021-06-12].

Anexo A. Deep learning

A.1. Redes neuronales artificiales

Los modelos de *machine learning* pueden clasificarse en supervisados o no supervisados. El *aprendizaje supervisado* se emplea para la *clasificación* de datos en función de diversas características, o para la *regresión*, si los datos han sido recopilados previamente, para predecir comportamientos futuros. Por otra parte, el *aprendizaje no supervisado* permite reconocer las diferencias entre datos y agruparlos en función de estas (técnica conocida como *clustering* o agrupación). Las técnicas de *machine learning* se emplean rutinariamente en múltiples aplicaciones, como asistentes virtuales en *smartphones*, vehículos autónomos, diagnóstico médico, bolsa, predicción de demanda eléctrica, o marketing, entre otras.

Como se explica en la Sección 4.1, las *redes neuronales artificiales* (NN) surgen como alternativa a la programación en computación convencional, y están inspiradas en los procesos que realiza el cerebro [108]. La investigación en NN ha aumentado notablemente desde el final de los años 80 y, actualmente, se aplica en numerosos campos de forma exitosa.

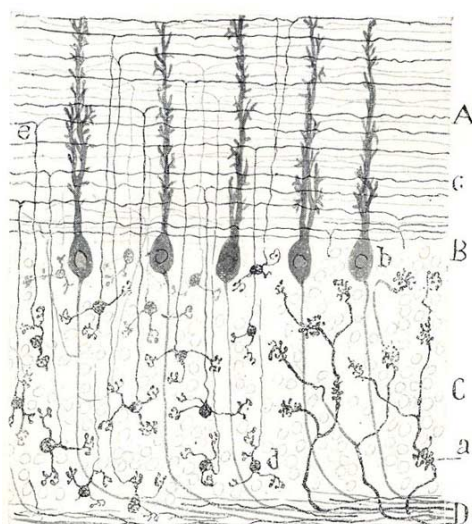
Las NN están formadas por neuronas artificiales: Unidades de procesamiento elemental que se interconectan entre sí dando lugar a la arquitectura de la red. Al igual que las que conforman el cerebro humano, las neuronas artificiales procesan la información, reconocen patrones, y aprenden a partir de datos o ejemplos proporcionados. La disposición de las neuronas en el cerebro humano se produce de forma autoorganizada, pudiendo darse en columnas, grupos neuronales, o capas, siendo esta última la organización que imitan las NN [94]. En la Figura A.1 se pueden observar las similitudes entre un modelo neuronal, dibujado por el científico Santiago Ramón y Cajal, y el modelo artificial.

A.1.1. Etapas de desarrollo de una red neuronal

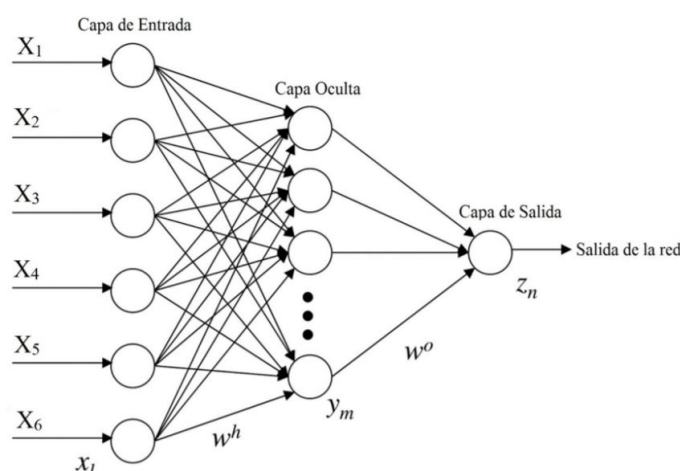
Para que las redes neuronales funcionen adecuadamente con respecto a los requisitos deseados, se siguen una serie de etapas de desarrollo en las cuales se buscan, recopilan y extraen conjuntos de datos o *datasets*, para posteriormente *entrenar* y *validar* la red, obteniendo así un algoritmo optimizado que, a través su prueba o *test*, permite obtener la salida esperada.

Recopilación de datos

La primera etapa es la recopilación de datos (*data collection*), que consiste en agrupar el suficiente número de ejemplos representativos que permita el correcto aprendizaje de la red.



(a) Neuronas situadas en capas intermedias de la corteza cerebral. Los cuerpos neuronales se unen mediante una red interconectada de procesos nerviosos [109, 110].



(b) Esquema general de una red neuronal artificial en el que se pueden visualizar las distintas capas en las que se organizan las neuronas [111].

Figura A.1: Redes neuronales: analogía entre el modelo biológico y el artificial.

En muchas ocasiones, se aplica un tratamiento previo a los datos con los que se trabaja, con intención de obtener mejores resultados y simplificar el sistema introduciendo un menor número de variables de entrada. Además, es muy recomendable realizar una normalización a estos datos; es decir, escalar sus valores para facilitar el aprendizaje.

Un número escaso de datos puede repercutir en un ajuste insuficiente del modelo al problema a resolver, generando resultados imprecisos o erróneos (fenómeno conocido como *underfitting* o subajuste). Por ejemplo, al entrenar una red neuronal para el reconocimiento de perros, si ha sido entrenada únicamente con una raza de perro, no asociaría otras razas de perro como el mismo animal con que se ha entrenado. Por el contrario, la introducción de datos en exceso en la red puede provocar *overfitting* o sobreajuste; es decir, no reconoce nuevos datos porque se ha ajustado a las particularidades de los datos previamente aportados. En el ejemplo mencionado de la red neuronal que reconoce perros, si es entrenada con demasiados ejemplos y se prueba con una imagen de un perro distinto, no lo reconocería como tal porque habría aprendido que el nuevo dato aportado no coincide con los datos empleados previamente en el entrenamiento [112].

Antes de comenzar la fase de entrenamiento, se realiza una selección de datos del *dataset* para dicha fase, con la particularidad de que van a ser únicamente empleados por la red en esta fase. Los datos restantes no seleccionados se emplearán en la fase posterior de *test*, y en ningún caso se mezclarán con los empleados para el entrenamiento.

Diseño de la red

Al enfocar un problema de *deep learning*, es importante definir la estructura del modelo a emplear, pues influye en la respuesta que la red proporciona, y en cómo la red se ajusta al problema. Para ello, hay que tener en cuenta las necesidades y requisitos que se tienen, las características

técnicas de la máquina empleada para su implementación, así como la organización y tamaño de los datos empleados, pues de ello dependerá el tiempo necesario para su computación.

Por ello, resulta fundamental realizar una esquematización previa de la arquitectura neuronal a emplear, pues facilita la comprensión del problema, el correcto tratamiento de los datos y, como consecuencia, la consecución de un modelo optimizado que produzca los resultados deseados. Además, el conocimiento del esquema de red que se utiliza puede ser de gran utilidad para futuras iteraciones de la estructura del modelo, permitiendo llevar un organizado control de versiones de este.

Entrenamiento

Durante la fase de aprendizaje o entrenamiento (en inglés, comúnmente denominada *train*), la red neuronal se adapta para resolver un problema mediante la introducción de ejemplos en la misma. Para ello, se parte de una red en la que se otorgan valores (pesos) iniciales a cada neurona. Existen numerosas técnicas, como el empleo de distribuciones para generar los pesos iniciales [113]. En esta fase, hay varias funciones que poseen un papel fundamental, como la *función de pérdida* o *loss*, que calcula el error o diferencia entre cada elemento predicho y su correspondiente real o *ground truth*. Conforme el entrenamiento avanza, los valores (pesos) que toman las neuronas se actualizan en cada paso (o *step*) mediante un *algoritmo de descenso de gradiente*. El objetivo del entrenamiento es que la función de pérdida alcance un mínimo global y, por tanto, el entrenamiento converja.

Además, los gradientes se multiplican por un *coeficiente de aprendizaje* o *learning rate*, que indica el grado de actualización de los pesos de las neuronas en cada paso. El coeficiente de aprendizaje ha de escogerse con precisión, pues con valores elevados, los pasos de actualización de los valores de las neuronas pueden ser demasiado grandes, pudiendo así pasar por alto el mínimo global que se busca. Por otra parte, emplear valores bajos puede ralentizar notablemente el entrenamiento, pudiendo incluso confundir un mínimo local con el mínimo absoluto. En general, al principio del entrenamiento, las neuronas de la red aprenden con mayor rapidez (elevado coeficiente de aprendizaje) puesto que todavía no conocen el problema.

En el entrenamiento, el modelo ha de pasar por todos los datos de *train* en cada iteración. El tiempo transcurrido en este proceso se conoce como *epoch*, término que se emplea con frecuencia para hacer referencia al número de veces que el modelo recibe todos los datos. Al principio de cada *epoch*, se separa aleatoriamente un subconjunto de los datos de *train* para, tras cada iteración, validar el modelo. El objetivo de realizar una validación en cada *epoch* es permitir la prueba del modelo, por lo que este no aprende ni actualiza sus pesos durante la validación, sino que únicamente se calcula el error para conocer su comportamiento. Si el error disminuye, se actualiza el modelo como mejor versión obtenida temporalmente, y si empeora, se mantienen los mejores valores obtenidos hasta el momento en iteraciones anteriores (ver Algoritmo 1).

Es importante remarcar que los datos de entrenamiento y validación se subdividen de forma aleatoria en cada *epoch*, y no se duplican en ningún caso, aunque son mezclados nuevamente en cada iteración. Además, para el *train* y la validación en cada *epoch*, los datos se agrupan en lotes o *batches* de menor tamaño, facilitando así su procesamiento, y se calculan las funciones de pérdida relativas a ambos procesos. A lo largo del proceso de entrenamiento, se comparan las funciones de pérdida obtenidas en cada *epoch*, con intención de minimizarlas. Para el valor de error mínimo alcanzado, se obtiene el modelo entrenado que mejor adaptado se encuentra al

problema a resolver.

Algoritmo 1: Algoritmo de entrenamiento y validación para cada *epoch*

```
Result: Modelo entrenado
separación de datos en batches
mejor_loss = inf
for epoch do
    loss =  $\emptyset$ 
    for batch do
        predicción = modelo(batch)
        error = calcular_error(predicción, ground truth(batch))
        loss = loss + error
    end
    loss = loss.mean()
    if train then
        modelo.optimización(loss)
    end
    if validation then
        if loss < mejor_loss then
            mejor_loss = loss
            guardar_modelo()
        end
    end
end
```

Test

Como se ha mencionado anteriormente, la fase de *test* se realiza con datos que no han sido previamente vistos por el modelo. Esta fase ocurre una vez la red ha sido entrenada y ha aprendido a resolver el problema planteado. Para ello, el modelo recibe datos de forma similar a la fase de entrenamiento, los procesa, y genera la salida para la cual ha sido entrenada. Finalmente, se realizan comparaciones de los resultados, pudiendo ser cuantitativas (mediante el cálculo de distintas métricas que indican de forma numérica, a través de números decimales, la relación entre la predicción obtenida y los datos *ground truth*), o cualitativas (mediante la comparación visual de los resultados).

A.2. Redes neuronales convolucionales

Las redes neuronales convolucionales (en inglés, *convolutional neural networks*, CNN) se emplean con imágenes, entendidas como matrices de píxeles [114]. Su nombre proviene del empleo de *convoluciones* o filtros que se aplican sobre la imagen. Estos filtros se utilizan en forma de ventana o *kernel* que, en función de su tamaño, cubre parcialmente la imagen y se desplaza por la misma en función de unos parámetros previamente definidos: el paso o *stride*, y el relleno o *padding*. Como se observa en la Figura A.2, el *stride* se refiere al número de píxeles que avanza el *kernel* en cada paso, y el *padding* indica el número de filas y columnas nulas que añade en los extremos de la matriz para evitar la pérdida de dimensionalidad tras hacer una convolución [115].

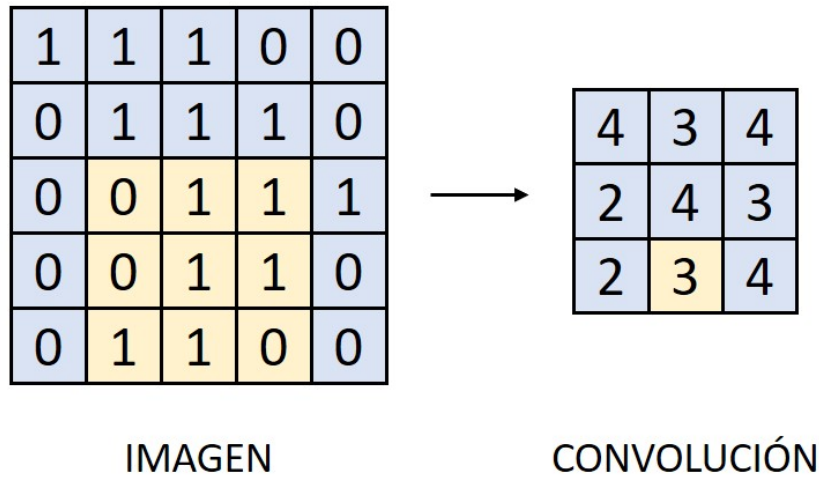


Figura A.2: Representación de aplicación de un filtro a una imagen, entendida como una matriz de píxeles [116].

Si bien el tamaño de la imagen queda representado por el número de filas y columnas de la matriz, las matrices no son bidimensionales, sino tridimensionales, pues cada elemento de la tercera dimensión representa un *canal* que contiene información de la imagen. Por ejemplo, una imagen RGB posee tres canales, uno por color, con los valores de rojo, verde y azul respectivamente para cada uno de los píxeles de la imagen. Por otra parte, una imagen en blanco y negro posee un único canal, con información del valor de cada píxel en la escala de grises. Además, al aplicar una convolución a una imagen, esta se aplica individualmente a cada uno de los canales de la imagen.

La finalidad de una red convolucional es la reducción del tamaño de las imágenes a una forma que sea más fácil de procesar, sin perder información crítica para obtener una buena predicción como resultado. Para ello, se aplican los filtros, que reducen sus dimensiones, aumentando el número de canales de las mismas, que dejan de tener un sentido físico o comprensible para el ser humano, pero sí pueden ser entendidas por un ordenador, siendo cada uno de esos canales, una “característica” distinta.

Las primeras capas convolucionales de una red capturan características de bajo nivel de la imagen, como los bordes, líneas, el color, etc. Pero, habitualmente, se combinan con otras funciones (o capas) que extraen características más sofisticadas progresivamente, formando una red que comprende completamente las imágenes, de una forma similar a la del ser humano.

Particularmente, en este proyecto, la red se organiza en *bloques convolucionales* (ver Figura A.3). En cada bloque, se realiza una convolución bidimensional (es decir, espacial) a cada canal de la imagen. Posteriormente, se normalizan los valores obtenidos en todo el conjunto o *batch* de imágenes con el que se está trabajando, pues esto agiliza y estabiliza el entrenamiento de la red; y finalmente, se utiliza una *función de activación*, que genera unos valores de salida a partir de los valores de entrada. En este caso, se emplea la función de activación denominada *unidad lineal rectificadora* (comúnmente conocida como ReLU, por sus siglas en inglés), que transforma los valores de entrada anulando los valores negativos y manteniendo los positivos sin ninguna alteración.

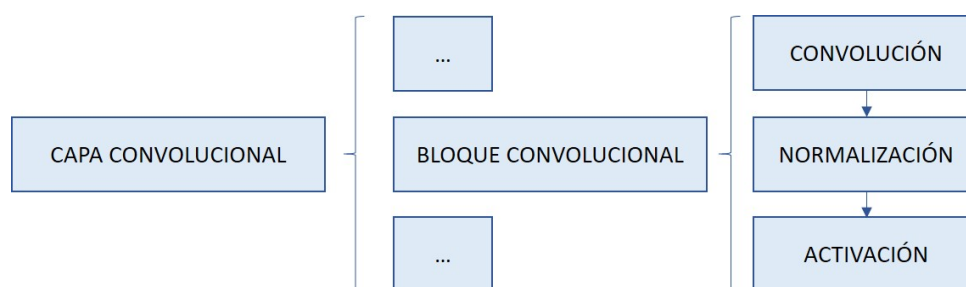


Figura A.3: Esquema de una capa convolucional. En este proyecto, formada por un único bloque convolucional: convolución, normalización y activación.

A.3. Modelo *encoder-decoder*

Un codificador-decodificador, comúnmente conocido por sus términos anglosajones *encoder-decoder*, es un modelo que sigue una estructura secuencial, y que es empleado ampliamente con imágenes y vídeos. El modelo *encoder-decoder* analiza las imágenes y extrae características descriptivas de las mismas a través de un proceso de codificación y decodificación, a continuación explicado.

La primera parte del proceso se da en el codificador o *encoder*, que transforma los datos proporcionados reduciendo sus dimensiones y aumentando el número de canales. Para ello, el *encoder* está formado por varios bloques convolucionales que, como se ha explicado en la Sección A.2, reducen el tamaño de las imágenes, aumentando el número de canales de las mismas. A la salida del *encoder*, los datos se traducen al formato requerido por la red, es decir, en un vector conocido como *vector latente*. Este proceso se realiza en el cuello de botella (*bottleneck*), formado por un bloque convolucional adicional que transforma la información de salida del *encoder* en el *vector latente* sin alterar sus dimensiones. Por tanto, el *vector latente* posee toda la información de los datos de entrada en forma vectorial, no comprensible para el ser humano [117].

Es entonces cuando el *decoder* entra en acción, pues se encarga de traducir la información condensada en el *vector latente* a unos datos de salida inteligibles. Para ello, realiza el proceso opuesto al *encoder*: Mediante varios bloques convolucionales, amplía la dimensionalidad de las imágenes y reduce el número de canales, obteniendo así la imagen deseada en la salida [118].

La arquitectura *encoder-decoder* es ampliamente utilizada en gran variedad de aplicaciones relacionadas con el procesamiento de imágenes. Por ello, se decide aplicar en este proyecto, introduciendo imágenes RGB de la vista delantera del vehículo en la entrada del *encoder*, obteniéndose imágenes en blanco y negro como mapas de saliencia en escala de grises.

Anexo B. Conjunto de datos de Dr(eye)ve

B.1. Datos empleados (*Dataset*)

El proyecto *Dr(eye)ve* [101], del grupo de investigación *AImageLab* [119] de la Universidad de Módena, surge con la intención de predecir la atención del conductor, mediante la estimación de aquellos elementos a los que es más probable que preste atención durante la conducción, así como mediante el reconocimiento de las zonas críticas de la escena en torno al vehículo.

Particularmente, a través del proyecto *Dr(eye)ve*, se dispone públicamente un conjunto de datos, o *dataset*, formado por 74 secuencias de vídeo de 5 minutos de duración cada una de ellas (un total de 555.000 fotogramas), en las que se ha grabado a ocho conductores diferentes en entornos diversos (ciudad, autopista, carreteras secundarias, etc.), y bajo condiciones variables de tráfico, luz (día, noche) y atmosféricas (soleado, nublado, lluvioso). Para evitar la acentuación de tendencias de fijación de la mirada de cada uno de los conductores en función de su forma de conducir, estos se han ido alternando a lo largo de toda la grabación (ver Tabla B.2).

Como se observa en la Figura B.1, el *dataset* del proyecto *Dr(eye)ve* proporciona múltiples datos relacionados con la atención del conductor: De cada fotograma, se proporciona la vista desde una cámara montada en el salpicadero del vehículo, el punto de vista desde los ojos del conductor, el mapa de saliencia visual obtenido mediante *eye tracking*, y la ruta georreferenciada del vehículo.

B.1.1. Datos empleados en el proyecto

Extracción de imágenes

Como se comenta en la Sección 4.2, en este proyecto se emplea una red neuronal con arquitectura *encoder-decoder*, que se caracteriza por el filtrado y la interpretación de las imágenes a través de múltiples convoluciones. Por ello, para la realización de este trabajo, se han tomado las secuencias de vídeo del *Dataset Dr(eye)ve* y se ha realizado una discretización en fotogramas: A partir de los 74 vídeos de 5 minutos de duración, grabados a una frecuencia de 25 fotogramas por segundo (fps), se obtienen 7500 fotogramas por cada archivo de vídeo (y, como se mencionaba previamente, un total de 555.000 fotogramas). Al tratarse de una cifra elevada de datos, se realiza una selección que permita la correcta comprensión secuencial del vídeo, sin que se pierda información entre fotogramas contiguos. Para ello, se estima un intervalo de selección cada 30 fotogramas, reduciendo así el número de datos a 250 fotogramas por secuencia (18500

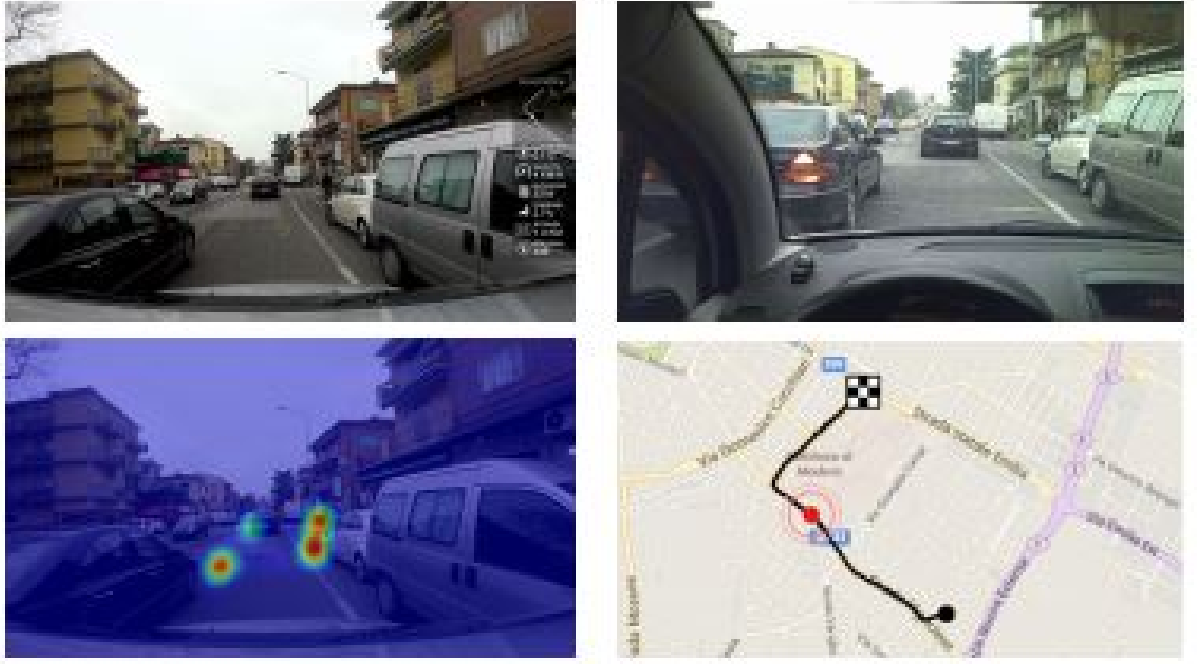


Figura B.1: Ejemplo del contenido del conjunto de datos de *Dr(eye)ve*. De izquierda a derecha: vista desde salpicadero, vista del conductor, mapa de saliencia y ruta georreferenciada [100].

# Vídeos	# Fotogramas	# Conductores	Condiciones atmosféricas	Iluminación	Información de la mirada	Metadatos	Cámara
74	555000	8	Soleado	Día	Coordenadas de mirada	GPS	Conductor (720p)
			Nublado	Atardecer	Mapa de saliencia	Velocidad del vehículo	Vehículo (1080p)
			LLuvioso	Noche	Dilatación de la pupila	Ruta	- - -

Tabla B.2: Resumen de las características del conjunto de datos de *Dr(eye)ve* [100].

fotogramas en total), siendo posteriormente empleados para el entrenamiento y test del modelo propuesto en este trabajo de fin de grado.

Análogamente, en el *dataset*, se proporciona un vídeo que recoge las coordenadas de fijación del conductor (saliencia *ground truth*) en cada fotograma de la secuencia de conducción. Por ello, se realiza el mismo proceso de extracción de fotogramas de saliencia, de tal forma que los fotogramas de saliencia e imágenes RGB que han sido separados se correspondan entre sí. Finalmente, para facilitar el trabajo con la red de forma ordenada, se realiza un emparejamiento los datos de tal forma que cada una de las imágenes RGB se encuentre asociada a su correspondiente de saliencia *ground truth*. Por tanto, cuando se trabaja con la red, los datos se encontrarán organizados por parejas en todo momento.

Transformaciones previas de la imagen

Originalmente, las imágenes extraídas tienen un tamaño de 1920 píxeles de anchura por 1080 píxeles de altura. Sin embargo, resulta ser una resolución grande que requeriría una la red neuronal con mayor número de parámetros, lo que dificultaría y ralentizaría su procesamiento. Por ello, antes de ser introducidas en el modelo, se realizan varias transformaciones que facilitarán el funcionamiento de la red. Además, la creación de un modelo que funcione adecuadamente con

B. Conjunto de datos de *Dr(eye)ve*

Transformación	Descripción
<i>Blur</i>	Difumina la imagen de entrada empleando un <i>kernel</i> de tamaño aleatorio.
<i>GaussNoise</i>	Agrega ruido Gaussiano con media y varianza definidas por el usuario. Permite su aplicación de forma independiente en cada canal.
<i>ColorJitter</i>	Modifica aleatoriamente brillo, contraste y saturación de una imagen.
<i>MotionBlur</i>	Aplica desenfoque con movimiento mediante un <i>kernel</i> de tamaño aleatorio.
<i>RandomShadow</i>	Simula sombras en el área de la imagen definida por el usuario.
<i>RandomSunFlare</i>	Simula rayos de sol en la imagen.

Tabla B.3: Transformaciones aplicadas a los datos en condiciones diurnas, obtenidas de la librería *Albumentations*.

imágenes de menor resolución puede ser de gran interés, pues indicaría que no se necesita instalar cámaras de alta resolución en el vehículo para el correcto funcionamiento de un sistema basado en el procedimiento planteado en este proyecto.

Por tanto, antes del paso de las imágenes por la red, se realizan diversas transformaciones empleando *OpenCV*¹ (una librería de visión por computador que goza de gran popularidad por su amplitud de aplicación en multitud de problemas). Los tratamientos aplicados a las imágenes incluyen su reescalado, que cambia el tamaño de imagen al deseado en la entrada de la red (256x128), y la normalización de los valores de cada uno de sus píxeles a números decimales entre 0 y 1 para evitar explosiones en los valores de los parámetros. En el caso de los mapas de saliencia *ground truth*, se realiza el mismo proceso.

B.1.2. Data augmentation

La técnica de aumento de datos o *Data Augmentation* consiste en el incremento del número de datos de un *dataset* con la finalidad de tener un tamaño mayor de muestra para entrenar y probar una red, y se trata de una técnica comúnmente empleada cuando se posee un *dataset* con un número insuficiente de datos. En este proyecto, se realiza *Data Augmentation* de los datos en condiciones diurnas, mediante el uso de la librería de Python *Albumentations*², que posee diversas transformaciones aplicables a las imágenes. Particularmente, en el proyecto se emplean seis transformaciones de gran utilidad para la conducción: difuminado (*blur*), ruido Gaussiano (*GaussNoise*), alteración de color (*ColorJitter*), desenfoque en movimiento (*MotionBlur*), sombras (*RandomShadow*), y luz solar directa (*RandomSunFlare*). La descripción de estas transformaciones se encuentra en la Tabla B.3.

Todas las transformaciones utilizadas poseen un parámetro $p \in [0, 1]$, que indica la probabilidad de que se aplique dicha transformación. En este proyecto, se define $p = 0.3$, para que este proceso desemboque en una media de dos transformaciones por imagen.

¹<https://opencv.org/>

²<https://albumentations.ai/docs/>